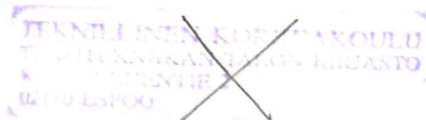




TEKNILLINEN KORKEAKOULU
Tietotekniikan osasto
Tietotekniikan koulutusohjelma

Mikko Haaramo

**ICONCLASS-LUOKITTELUJÄRJESTELMÄN
ONTOLOGISOINTI JA SOVELTAMINEN**



Diplomityö, joka on jätetty opinnäytteenä tarkistettavaksi
diplomi-insinöörin tutkintoa varten Espoossa 17.11.2006

Valvoja professori Petri Vuorimaa

Ohjaaja professori Eero Hyvönen

TEKNILLINEN KORKEAKOULU Tietotekniikan osasto		DIPLOMITYÖN TIIVISTELMÄ	
Tekijä Mikko Haaramo		Päiväys 17.11.2006	
		Sivumäärä 67 + 3	
Työn nimi Iconclass-luokittelujärjestelmän ontologisointi ja soveltaminen			
Professuuri Vuorovaikutteinen digitaalinen media		Koodi T-111	
Työn valvoja Professori Petri Vuorimaa			
Työn ohjaaja Professori Eero Hyvönen			
<p>Tämän diplomityön tavoitteena oli selvittää, miten kuvataideteosten sisällönkuvailuun käytettävä Iconclass-järjestelmä voidaan ontologisoida ja yhdistää Yleiseen suomalaiseen ontologiaan (YSO) ja sitä kautta muihin ontologioihin. Diplomityö tehtiin osana Semanttisen laskennan tutkimusryhmän (SeCo) Suomalaiset semanttisen webin ontologiat (FinnONTO) -projektia. Projektissa kehitetään keskeisiä kansallisia ontologioita, ontologioiden kehittämisen ja käytön palveluita, sekä pilottisovelluksina semanttisia portaaleja.</p> <p>Työssä verrattiin Iconclass-järjestelmän rakennetta muiden ontologioiden ja tesaurusten rakenteeseen. Järjestelmän todettiin olevan hierarkkiselta perusrakenteeltaan näiden kaltainen, mutta sisältävän myös semantiikaltaan monimutkaisempia laajennuksia. Lisäksi tesauruksista poiketen käsitteiden yksilöintiin käytetään notaatioita eli kieliriippumattomia luokittelukoodeja. Ontologisoinnin kannalta haasteeksi todettiin erityisesti järjestelmän laajennusten semanttisen rakenteen esittäminen.</p> <p>Iconclass-järjestelmän rakenteen analyysin perusteella laadittiin malli ICONCLASS-ontologialle laajentamalla tesaurusten perusrakenteen esittämiseen tarkoitettua SKOS Core -mallia. Malli mahdollistaa luokittelujärjestelmän notaatiohierarkian ja tähän liittyvän avain-laajennuksen kuvaamisen. Ontologia yhdistettiin YSO:aan etsimällä käsitevastineita Iconclass-järjestelmän luokkiin liittyvien avainsanojen ja YSO:n käsitteiden välillä.</p> <p>Ontologian soveltumista konesemanttiseen päättelyyn kokeiltiin käyttämällä sitä Valtion taidemuseon maalausten liittämiseen erityyppisiä kulttuurisisältöjä yhdistävään Kulttuurisampo-portaaliin. Lisäksi selvitettiin myös mahdollisuuksia ontologiaan perustuvaan teosten luokittelun avustamiseen. Konesemanttinen päättely avaa uusia mahdollisuuksia Iconclass-järjestelmän hyödyntämiseen, mutta jatkokehityksessä on syytä selvittää myös miten koneellinen päättely vastaa käyttäjien tarpeisiin.</p>			
Avainsanat Iconclass, luokittelujärjestelmä, ontologia, semanttinen web			

HELSINKI UNIVERSITY OF TECHNOLOGY ABSTRACT OF MASTER'S THESIS Department of Computer Science and Engineering	
Author Mikko Haaramo	Date 17.11.2006
	Pages 67 + 3
Title of thesis Ontologization and application of Iconclass classification system	
Professorship Interactive Digital Media	Professorship Code T-111
Supervisor Professor Petri Vuorimaa	
Instructor Professor Eero Hyvönen	
<p>The aim of this Master's thesis was to study the ontologization of Iconclass system and mapping it with the General Finnish Upper Ontology (YSO) and other ontologies. Iconclass is a subject specific classification system for describing contents of pictures. This thesis was written as part of National Semantic Web Ontology Project in Finland (FinnONTO) carried out by the Semantic Computing Research Group (SeCo). The goal of this project is to develop national core ontologies, ontology services and pilot application.</p> <p>The structure of the Iconclass system was compared with other ontologies and thesauri. These share similar hierarchic basic structure but Iconclass system has also various auxiliaries with complex semantics. Unlike thesauri Iconclass uses language independent classification codes i.e. notations to identify concepts. Representing the semantics of Iconclass auxiliaries was identified as one of the major challenges of the ontologization.</p> <p>After the analysis of Iconclass system a model for ICONCLASS ontology was created by extending the SKOS Core model, which is a ontology model for representing thesauri. The model represents the notation hierarchy of Iconclass system and its key hierarchies. The ontology was mapped with YSO by matching keywords from Iconclass index with labels of YSO concepts.</p> <p>The ontology was used to link the collections of the Finnish National Gallery with CultureSampo portal which combines heterogenous cultural content. The feasibility of building an ontology-based classifier helper tool was also studied. Semantic reasoning enables new possibilities for utilization of the Iconclass system, however, further research is needed to evaluate how these applications match users needs.</p>	
Keywords Iconclass, classification system, ontology, semantic web	

ALKUSANAT

Tämä diplomityö on tehty Teknillisen korkeakoulun Viestintätekniikan laboratoriossa Semanttisen laskennan tutkimusryhmässä. Haluan kiittää työni ohjaajaa professori Eero Hyvöstä mielenkiintoisen aiheen tarjoamisesta. Kiitokset myös työni valvojalle professori Petri Vuorimaalle.

Kiitos tutkimusryhmän henkilökunnalle kannustavasta työilmapiiristä. Valtion taidemuseota kiitän yhteistyöstä projektin aikana sekä aineiston tarjoamisesta tutkimuskäyttöön.

Kiitokset myös kotiväelle tuesta ja kannustuksesta.

Helsingin Aurinkolahdessa 17.11.2006

Mikko Haaramo

SISÄLLYSLUETTELO

1	Johdanto	1
1.1	Taustaa	1
1.2	Tavoitteet	1
1.3	Diplomityön rakenne	2
2	Ontologioista	4
2.1	Ontologian määritelmiä	4
2.2	Ontologiat semanttisessa webissä	5
2.3	Semanttiset suhteet tesauroyksissa	6
2.4	Tesauroksesta ontologiaksi	8
2.5	Yhteenvedo	9
3	Iconclass-järjestelmä	10
3.1	Järjestelmän kehitys ja käyttö	10
3.2	Iconclass-järjestelmän osat	11
3.3	Luokittelujärjestelmän perusrakenne	12
3.3.1	Notaatiot ja määritteet	12
3.3.2	Hierarkkinen rakenne	12
3.3.3	Semanttiset suhteet	14
3.4	Luokittelujärjestelmän laajennukset	15
3.4.1	Nimi sulkeissa	15
3.4.2	Kaksoiskirjaimet	17
3.4.3	Rakenteellinen numerointi	17
3.4.4	Avaimet	18
3.5	Aakkosellinen indeksi	20
3.6	Semanttiset sovellukset	21
3.7	Yhteenvedo	22
4	Haasteet konesemantiikalle	23
4.1	Luokittelujärjestelmän rakenteen kuvaaminen	23
4.1.1	Käsitteiden nimikkeet	23
4.1.2	Semanttiset suhteet	25
4.1.3	Avain-laajennuksen rakenne	27
4.2	Yksilöivän tunnisteiden muodostaminen ontologian käsitteille	28
4.2.1	Notaatioiden syntaksi	28
4.2.2	URI-tunnisteiden syntaksi	28
4.2.3	Rajoitteet notaation esittämiseksi URI-tunnisteena	30
4.2.4	URI-tunnisteiden sarjallistaminen XML-muotoon	31
4.2.5	Rajoitteet notaation esittämiseksi XML-elementin nimenä	31
4.2.6	ICONCLASS-ontologian URI-tunnisteet	32
4.3	Luokittelun merkityksen esittäminen	33
4.3.1	Käsitevastaavuudet	33
4.3.2	Tapahtumakuvaus	34
4.4	Yhteenvedo	35

5	ICONCLASS-ontologia	36
5.1	ICONCLASS-ontologian malli	36
5.1.1	SKOS Core -mallin puutteet	36
5.1.2	ICONCLASS-laajennus	36
5.2	Muunnos ontologiaksi	37
5.2.1	Vastaavuuksien määrittely	37
5.2.2	Luokittelujärjestelmän laajennusten käsittely	39
5.2.3	Muunnoksen toteutus	40
5.3	Iconclass-ontologian yhdistäminen muihin ontologioihin	41
6	Sovelluskohteita	43
6.1	KulttuuriSampo -portaali	43
6.1.1	Iconclass-aineisto	43
6.1.2	Muunnosmenetelmä	43
6.1.3	Tulosten arviointi	48
6.2	Luokittelijan apuväline ICONCLASS-ONKI	50
6.2.1	Oikeiden notaatioiden löytäminen	51
6.2.2	Olennessen sisällön esille nostaminen	53
6.2.3	Luokittelun tehokkuus	53
6.2.4	Hierarkkisen rakenteen hankaluus	53
6.2.5	Toteutusvaiheet	54
7	Arviointi	55
7.1	SKOS Core -malli	55
7.2	ICONCLASS-ontologian malli	55
7.3	Ontologian sovellukset	56
7.4	Lähdeaineiston puutteet	57
8	Johtopäätökset	58
8.1	Vastaukset tutkimuskysymyksiin	58
8.2	Jatkotutkimuskohteita	61
8.3	Tulevaisuus	62
	Lähdeluettelo	63
	Liitteet	
	Liite 1. Iconclass-laajennus SKOS Core -malliin	
	Liite 2. Maalauksen Kullervon kirous Iconclass-annotaatio	

LYHENNELUETTELO

AAT	<i>Art & Architecture Thesaurus.</i> Getty-säätiön kehittämä tesauros taiteen ja arkkitehtuurin luokitteluun.
CIDOC CRM .	<i>CIDOC Conceptual Reference Model.</i> Museoiden kokoelmien kuvailuun tarkoitettu ontologia.
KNAW	<i>Koninklijke Nederlandse Akademie van Wetenschappen.</i> Alankomaiden kuninkaallinen tiedeakatemia. Iconclass-järjestelmän hallinnosta aiemmin vastannut organisaatio.
NLP	<i>Natural Language Processing.</i> Luonnollisen kielen koneellista tulkitsemista tutkiva tekoälytutkimuksen ja kielitieteen osa-alue.
OWL	<i>Web Ontology Language.</i> W3C:n suositus semanttisen webin ontologiakieleksi.
RDF	<i>Resource Description Framework.</i> W3C:n suositus metatiedon kuvaamiseen webissä.
RDFS	<i>RDF Schema.</i> RDF-kielellä määritelty sanastonkuvauskieli hierarkkisten sanastojen esittämiseen.
RKD	<i>Rijksbureau voor Kunsthistorische Documentatie.</i> Alankomaiden taidehistoriallinen instituutti. Iconclass-järjestelmän hallinnosta ja kehitystyöstä vastaava organisaatio.
SECO	<i>Semanttisen laskennan tutkimusryhmä.</i>
SKOS	<i>Simple Knowledge Organization Framework.</i> Kehitteillä oleva RDFS-kieltä laajentava kieli tesaurusten ja luokittelujärjestelmien esittämiseen semanttisessa webissä.
TEKES	<i>Teknologian ja innovaatioiden kehittämiskeskus.</i>
URI	<i>Uniform Resource Identifier.</i> Yksikäsitteinen merkkijonotunniste, jolla resursseihin viitataan webissä.
US-ASCII	<i>United States of America Standard Code for Information Interchange.</i> Tietokoneiden merkistöstandardi, joka sisältää englannin kielen kirjaimet ja numerot sekä välimerkkejä ja ohjauskoodeja.
W3C	<i>World Wide Web Consortium.</i> Webin teknologioita kehittävä ja standardoiva kansainvälinen järjestö.
WWW	<i>World Wide Web.</i>
XML	<i>Extensible Markup Language.</i> W3C:n suositus yleiskäyttöiseksi ja laajennettavaksi rakenteisen tiedon merkintäkieleksi.
XSL	<i>Extensible Stylesheet Language.</i> XML-muotoisen aineiston muotoiluun ja esittämiseen tarkoitettu kieli.
XSLT	<i>XSL Transformations.</i> Kuvauskieli XML-tiedostojen muunnoksiin.
YSO	<i>Yleinen suomalainen ontologia.</i> Yleisen suomalaisen asiasanaston (YSA) pohjalta kehitettävä suomalainen yleisontologia.

1 JOHDANTO

1.1 Taustaa

Iconclass on kansainvälisesti käytetty aiheenmukainen luokittelujärjestelmä taide-
teosten sisällönkuvailuun. Semanttisessa webissä luokittelujärjestelmien käyttäjiä
eivät enää ole vain ihmiset, vaan niitä hyödynnetään myös koneellisessa päättelys-
sä. Ontologisointi tarkoittaa luokittelujärjestelmän muuntamista ontologiaksi, tie-
tokoneella käsiteltävissä olevaksi käsitemalliksi. Tämä antaa mahdollisuuden hyö-
dyntää semanttista päättelyä Iconclass-luokittelulla kuvailtujen teosten haussa sekä
kehittää ontologiaperusteisia menetelmiä kuvasisältöjen käsittelyyn.

Tämä diplomityö on tehty Teknillisen korkeakoulun Viestintätekniikan laborato-
riossa osana Semanttisen laskennan tutkimusryhmän (SeCo) Suomalaiset semant-
tisen webin ontologiat (FinnONTO) -projektia, joka kuuluu Tekesin FENIX-tekno-
logiaohjelmaan. FinnONTO-projektissa kehitetään muun muassa keskeisiä kansal-
lisia ontologioita sekä ontologioiden kehittämisen ja käytön palveluita. (Hyvönen,
2005a)

Diplomityössä esitetään menetelmä Iconclass-luokittelujärjestelmän ontologisoimi-
seksi sekä tarkastellaan ontologian käyttöä Iconclass-järjestelmän tyypillisimmissä
sovelluskohteissa teosten haussa ja luokittelussa.

1.2 Tavoitteet

Diplomityön tavoitteena on mahdollistaa kuvataideteosten sisällönkuvailuun käy-
tettävän Iconclass-järjestelmän konesemanttinen käyttö ja yhteentoimivuus kone-
semantiikkaa hyödyntävien sovellusten kuten KulttuuriSammon kanssa (Hyvönen
et al., 2006). Työssä selvitetään, miten Iconclass-luokittelujärjestelmä voidaan onto-
logisoida ja yhdistää Yleiseen suomalaiseen ontologiaan (YSO) ja sitä kautta muihin
ontologioihin. Työssä tarkastellaan myös, miten ICONCLASS-ontologian ja koneel-
lisen päättelyn avulla voidaan tukea käyttäjän toimintaa Iconclass-järjestelmän ylei-
simmissä sovelluskohteissa.

Työn tavoite on jaettu viiteen osakysymykseen. Kolme ensimmäistä kysymystä liittyvät ontologisoinnin toteutukseen ja kaksi jälkimmäistä ontologian yhdistämiseen muihin ontologioihin sekä ontologian hyödyntämiseen konesemanttisissa sovelluksissa.

1. Millainen Iconclass-järjestelmän rakenne on verrattuna tesaurusten ja ontologioiden rakenteeseen?

Tähän kysymykseen vastataan kirjallisuuden pohjalta selvittämällä mistä osista Iconclass-järjestelmä koostuu, miten luokittelujärjestelmä toimii ja millaisia rakenteita se sisältää.

2. Mitä ongelmia Iconclass-järjestelmässä on konesemantiikan kannalta?

Kysymykseen vastataan tarkastelemalla Iconclass-järjestelmän rakennetta ja sen tyyppisiä sovelluskohteita ja vertailemalla näitä semanttisen webin arkkitehtuurimallin ontologialle asettamiin vaatimuksiin.

3. Miten Iconclass-järjestelmän ontologisointi toteutetaan?

Kahteen edelliseen kohtaan perustuen esitellään malli Iconclass-ontologialle ja menetelmä mallia noudattavan ontologian muodostamiseen Iconclass-järjestelmästä.

4. Miten ICONCLASS-ontologian käsitteiden merkityksiä voidaan kuvata?

Kysymykseen vastataan esittämällä ontologioiden yhdistämiseen soveltuvia menetelmiä ja arvioimalla niiden soveltuvuutta ICONCLASS- ja YSO-ontologioiden yhdistämiseen.

5. Miten ICONCLASS-ontologia soveltuu koneelliseen päättelyyn?

Kysymykseen vastataan tarkastelemalla esimerkkien avulla koneellisen päättelyn etuja Iconclass-järjestelmän yleisimmissä sovelluskohteissa.

1.3 Diplomityön rakenne

Luvussa 2 käsitellään ontologioita ja niiden esittämiseen soveltuvia semanttisen webin teknologioita sekä tesaurusten ja luokitusjärjestelmien yleisimpiä semanttisia

suhteita ja näiden muuntamista ontologiaksi. Luvussa 3 esitellään Iconclass-järjestelmän osat, luokittelujärjestelmän rakenne laajennuksineen ja sen käyttö teosten annotoinnissa. Lopuksi luodaan katsaus Iconclass-järjestelmän käyttöön semanttisissa sovelluksissa.

Luvussa 4 käsitellään Iconclass-järjestelmän konesemanttisen käsittelyn haasteita, jotka liittyvät luokittelujärjestelmän ontologisointiin ja yhdistämiseen muihin ontologioihin. Luvussa 5 esitellään ICONCLASS-ontologian malli, jonka mukaisesti luokittelujärjestelmän ontologisointi tässä työssä toteutettiin.

ICONCLASS-ontologian käyttöä teosten semanttiseen hakuun sekä teosten luokittelun avustamista konesemanttisen päättelyn avulla käsitellään luvussa 6. Luvussa 7 arvioidaan SKOS Core -mallin soveltuvuutta luokittelujärjestelmien esittämiseen sekä tässä työssä esitetyn mallin mukaisen ICONCLASS-ontologian soveltumista Iconclass-järjestelmän yleisimpiin käyttökohteisiin. Luvussa 8 esitetään johtopäätöksiä vastaukset tutkimuskysymyksiin sekä esitetään jatkotutkimuskohteita ja työssä käsiteltyjen esimerkkisovellusten tulevaisuudennäkymiä.

2 ONTOLOGIOISTA

Tässä luvussa selvitetään, mitä ontologiat ovat, miten niitä esitetään semanttisessa webissä ja miten tesaureista voidaan muodostaa ontologioita. Ensiksi tutkitaan ontologian määritelmiä. Tämän jälkeen käsitellään ontologioiden koneelliseen esittämiseen soveltuvia kieliä keskittyen erityisesti tesaureista muodostettuihin ontologioihin. Lopuksi tarkastellaan tesaauruksen semanttisen rakenteen analysointia ja tesaauruksen ontologisointiin soveltuvia menetelmiä.

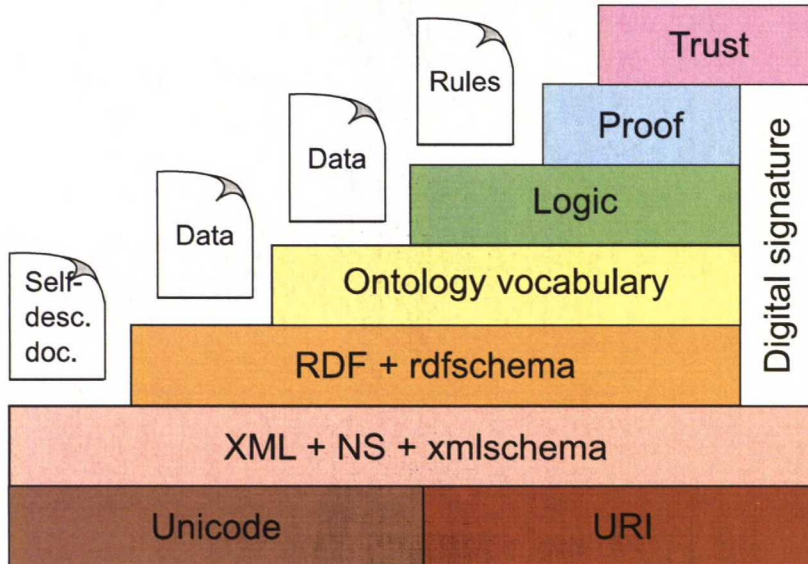
2.1 Ontologian määritelmiä

Alkuperäisessä merkityksessään sana ontologia liittyy filosofiaan, jossa sillä tarkoitetaan oppia olevaisesta ja sen rakenteesta (Lowe, 2005). Tekoälytutkimukseen ja sen kautta myös laajemmin tietojenkäsittelytieteeseen ontologiat tulivat 1990-luvulla (Fensel et al., 2003, s. 11). Samalla sana on saanut uuden merkityksen. Gruber (1993) määrittelee ontologian ”formaaliksi ja eksplisiittiseksi määrittelyksi yhteiselle käsitteistölle”. Käsitteistö on tiettyä ilmiötä kuvaava abstrakti malli, joka yksilöi ilmiön kannalta olennaiset käsitteet. Formaalius tarkoittaa, että ontologia on esitettävä koneellisesti käsiteltävässä muodossa. Eksplisiittisyys edellyttää käsitteiden ja niiden käytön rajoitteiden tarkkaa määrittelyä. Yhteisyys korostaa, että ontologian tehtävänä on kuvata aihealuetta koskevaa jaettua tietämystä. (Fensel, 2004, s. 3)

Myös Poole et al. (1998, s. 11) esittää samankaltaisen määritelmän ontologialle. Heidän mukaansa ”ontologia on sopimus, joka määrittelee millaisia yksilöitä ja suhteita tiettyyn aihealueeseen kuuluu”. Tämän määritelmän perustana on ajatus, että maailmaa voidaan kuvata yksilöiden ja näiden välisten suhteiden avulla. Hendlerin (2001) määritelmän mukaan ontologiaan voi liittyä myös yksinkertaisia loogisia sääntöjä. Hänen määritelmässään ontologia koostuu aihealueen termeistä sisältäen sanaston, näiden välisistä merkityssuhteista ja yksinkertaisista päättelysäännöistä. Monimutkaisemmat loogiset säännöt Hendler erottaa ontologiasta omaksi kokonaisuudekseen.

Kaikille näille määritelmille on yhteistä, että ontologiassa on yksilöitä, käsitteitä ja näiden välisiä merkityssuhteita. Nämä on määriteltävä yksikäsitteisesti ja tarkasti,

jotta ontologian käyttäjät voivat olla varmoja niiden merkityksestä ja jotta ontologiaa olisi mahdollista käsitellä myös tietokoneen avulla. Käsitteisiin liittyvien rajoitteiden määrittelyssä voidaan käyttää apuna tarvittaessa myös loogisia päättelysääntöjä.



Kuva 2.1: Semanttisen webin arkkitehtuuri (Berners-Lee, 2000)

2.2 Ontologiat semanttisessa webissä

Semanttisissa sovelluksissa ontologioiden määrittelemiä yhteisiä käsitteistöjä käytetään tietomallina agenttiohjelmistoille, joiden avulla voidaan esimerkiksi tehostaa tiedonhakua tai suorittaa ontologiaan perustuvaa koneellista päättelyä, kuten esittää käyttäjälle ontologiaan perustuvia suosituksia aineistosta (Berners-Lee et al., 2001). Tällaisten sovellusten toteuttamiseksi tarvitaan kieli, jolla ontologia voidaan esittää koneellisesti käsiteltävässä muodossa.

Semanttisissa sovelluksissa käytettävät ontologioiden esityskielet perustuvat resursien kuvaamiseen käytettävään RDF-kieleen. (Fensel, 2004, s. 21–46) Sanastonkuvauskieli RDF Schema (RDFS) määrittää ontologioiden perusrakenteen mahdollistamalla luokka- ja ominaisuushierarkioiden esittämisen (Brickley & Guha, 2004). Web Ontology Language (OWL) on W3C:n suositus ontologioiden esityskieleksi semanttisessa webissä. Se laajentaa RDFS:aa ja antaa mahdollisuuden kuvata semanttisesti rikkaita monimutkaisia suhteita ja rajoitteita sisältäviä ontologioita (McGuinness & van Harmelen, 2004). Kuva 2.1 esittää semanttisen webin arkkitehtuurin ja

siinä käytettävien kielten väliset suhteet. OWL:a on kritisoitu sovelluskohteiden tarpeisiin nähden liian laajaksi ja erityisosaamista vaativaksi (Miles et al., 2005). Simple Knowledge Organization Framework (SKOS) pyrkii vastaamaan tarpeeseen yksinkertaiselle, mutta silti riittävän tehokkaalle kielelle käsittemallien, kuten tesaurusten ja luokittelujärjestelmien esittämiseen semanttisessa webissä (Miles et al., 2005; Miles & Brickley, 2005). SKOS koostuu kolmesta osasta:

SKOS Core on käsitekeskeinen malli tesaurusten ja luokittelujärjestelmien rakenteen kuvaamiseen. Se laajentaa RDFS:aa määrittelemällä joukon luokkia ja ominaisuuksia, joilla käsitehierarkioita voidaan esittää semanttisessa webissä. (Miles & Brickley, 2005) SKOS Core -malli on tällä hetkellä luonnosvaiheessa, ja sen kehitystyötä W3C:n suositukseksi jatketaan (Miles, 2006).

SKOS Mapping on luonnos tesaurusten välisten käsitevastaavuuksien esittämiseen soveltuvasta mallista. SKOS Mappingin kehitystyö on keskeytynyt ja odottaa SKOS Core -mallin valmistumista. (Miles, 2005c)

SKOS Extensions on luonnos SKOS Core -mallin laajennuksesta, joka mahdollistaa semanttisten suhteiden tyyppin tarkemman kuvaamisen. Myös tämän osan kehitys jatkuu vasta SKOS Core -mallin valmistuttua. (Miles, 2005b)

2.3 Semanttiset suhteet tesauruksissa

Tesaurun rakenteen analysointiin liittyy olennaisena osana sen semanttisen rakenteen selvittäminen, sillä kuvattaessa tesaurusta ontologiamallin mukaiseen muotoon on tesaurun sisältämille semanttisille suhteille määriteltävä vastineet ontologiassa. Tässä luvussa tarkastellaan yleisimpiä tesauruksissa esiintyviä semanttisia suhteita sekä niiden esittämistä SKOS Core -mallissa.

Aitchison et al. (2000, ref. Hyvönen, 2005b, s. 4) jakaa tesauruksissa esiintyvät semanttiset suhteet kolmeen ryhmään: ekvivalenssisuhteisiin, hierarkkisiin suhteisiin ja assosiatiivisiin suhteisiin. Tässä jaottelussa hierarkkiset suhteet kattavat sekä ala–yläkäsitemallin että osa–kokonaisuus-suhteen. Suomenkielisessä kirjallisuudessa hierarkkinen suhde viittaa yleensä vain ala–yläkäsitemallin suhteeseen ja osa–kokonaisuus-

Taulukko 2.1: Tärkeimpiä semanttisia suhteita tesauruksissa ja niitä vastaavat ominaisuudet SKOS Core -mallissa (Hyvönen, 2005b; Miles & Brickley, 2005)

Suhde englanniksi	Suhde suomeksi	SKOS Core	Merkitys
BT Broader term	LT	skos:broader	Laajempi termi, hierarkkinen suhde, käänteissuhde ST
NT Narrower term	ST	skos:narrower	Suppeampi termi, hierarkkinen suhde, käänteissuhde LT
RT Related term	RT	skos:related	Rinnakkaistermi, assosiatiiivinen suhde
USE Use	KÄYTÄ	skos:prefLabel	Ilmaisee suositeltavan synonyymien, käänteissuhde KORVAA
UF Used for	KORVAA	skos:altLabel	Ilmaisee ei-suositeltavan synonyymien, käänteissuhde KÄYTÄ
SN Scope note	HUOMAUTUS	skos:scopeNote	Luonnehtii termin merkitystä ja käyttöä

suhteista käytetään nimitystä koostumussuhde (Nykänen & Kalliokuusi, 1999). Koska SKOS Core -malli ja Iconclass-järjestelmä eivät erottele ala-yläkäsité- ja osakokonaisuus-suhteita toisistaan, käytetään tässä työssä Aitchisonin esittämää jaottelua.

Hyvönen (2005b, s. 4) esittää listan tärkeimmistä semanttisista suhteista tesauruksissa. Taulukko 2.1 esittää tämän listan sekä suhteita vastaavat ominaisuudet SKOS Core -mallissa. Koska SKOS Core -malli ei kuvaa termejä vaan käsitteitä, voidaan KÄYTÄ- ja KORVAA-suhteilla ilmaistut samaan käsitteeseen liittyvät suositeltavat ja ei-suositeltavat synonyymit esittää käsitteen ensisijaisena ja vaihtoehtoisina nimikkeinä.

2.4 Tesauruksesta ontologiaksi

Tesaurusten muuntamiseen ontologiaksi on kehitetty useita menetelmiä, joista osa soveltuu yksittäisen tesauruksen muunnokseen ja osa on yleiskäyttöisiä. van Assem et al. (2006) vertailee erilaisia muunnosmenetelmiä tavoitteenaan kehittää menetelmä, joka soveltuu tesauruksen muuntamiseen SKOS Core -mallin mukaiseen muotoon.

Oman menetelmänsä pohjaksi van Assem et al. valitsee Miles et al. (2004) kehittämän kolmivaiheisen menetelmän, jonka tavoitteena on SKOS Core -mallin avulla tukea tesaurusten välistä yhteentoimivuutta säilyttäen samalla kaikki alkuperäiseen tesaurukseen sisältyvä informaatio. Menetelmän vaiheet ovat:

- 1. RDF-muodossa olevan tesauruksen tuottaminen
- 2. virheiden tarkistus ja validointi
- 3. julkaiseminen webissä

Menetelmän ensimmäistä vaihetta van Assem et al. (2006) täsmentää jakamalla sen taulukossa 2.2 esitettyihin osavaiheisiin, jotka määrittelevät tarkemmin, miten SKOS Core -mallin mukainen RDF-kuvaus tuotetaan tesauruksesta.

Taulukko 2.2: Vaiheen 1 osavaiheet ja tehtävät (van Assem et al., 2006) (suomennettu)

Osavaihe	Tehtävät	Tulos
(A) tesauruksen analysointi	digitaalisen tallennusmuodon ja dokumentaation analysointi	luettelo tietoalkioista ja rajoitteista, lista tesaurusten ominaisuuksista
(B) kuvaus SKOS-malliin	vastaavuuksien määrittely tietoalkioille SKOS-skeemassa	taulukko tietoalkioiden vastineista skeemassa
(C) muunnosohjelma	algoritmin kehittäminen	muunnosohjelma

Tämä menetelmä valittiin myös Iconclass-järjestelmän ontologisoinnin pohjaksi. Kriteerejä valinnalle olivat menetelmän yleiskäyttöisyys, joka mahdollistaa sen soveltamisen erilaisille lähdeaineistoille, muunnoksen osavaiheiden tarkka kuvaus sekä

aineiston sisältämän informaation säilyminen muunnoksessa. Huomiota kiinnitettiin myös siihen, että menetelmä muuntaa aineiston SKOS Core -mallin mukaiseen muotoon, mikä oli alustavasti valittu ontologian esitystavaksi.

2.5 Yhteenveto

Ontologiat ovat käsitemalleja, jotka kuvaavat tietyn aihealueen käsitteitä ja näiden välisiä suhteita. Semanttisessa webissä ontologioita käytetään käsitteiden merkitysten määrittelyyn ja niiden välisten riippuvuuksien esittämiseen koneellisessa päätelyssä. Useilla aihealueilla ontologioita on muodostettu erilaisten tesaurusten pohjalta. Useimpien tesaurusten perusrakenne muodostuu käsitteiden välisistä hierarkkisista ja assosiatiivisista suhteista. Tesaauruksen muuntamiseen ontologiaksi on kehitetty sekä yleiskäyttöisiä että yksittäisen sanastojen muuntamiseen soveltuvia menetelmiä. Muunnoksessa analysoidaan tesaauruksen semanttinen rakenne ja kuvataan se ontologian esityskielelle.

3 ICONCLASS-JÄRJESTELMÄ

Tässä luvussa esitellään Iconclass-järjestelmä. Ensin käsitellään järjestelmän historiaa, käyttökohteita ja käyttäjäryhmiä. Tämän jälkeen tarkastellaan itse järjestelmää, sen osia sekä järjestelmän semanttista rakennetta. Lopuksi luodaan katsaus aiempiin semanttisiin sovelluksiin, joissa Iconclass-järjestelmää on hyödynnetty.

Iconclass-järjestelmän osista käytetään pääsääntöisesti samoja suomenkielisiä nimityksiä kuin mitä Anttosen (2003) käännösmonisteessa ja Lüscherin (2004) Pro gradu-tutkielmassa käytetään. Koska järjestelmään liittyvän suomenkielisen terminologian ei voi katsoa olevan vielä vakiintunutta, on vastaavat englanninkieliset käsitteet mainittu sulkeissa.

3.1 Järjestelmän kehitys ja käyttö

Iconclass on aiheenmukainen hierarkkinen luokittelujärjestelmä kuvataideteosten sisällönkuvailuun. Sen on kehittänyt hollantilainen taidehistorian professori Henri van de Waal. Järjestelmän kehitystyö alkoi 1947 ja luokittelujärjestelmä julkaistiin painettuna versiona vuosina 1973–1985. (van den Berg, 1995; Lüscher, 2004, s. 15)

Nykyisin Iconclass-järjestelmän kehitystyöstä ja hallinnosta vastaa Alankomaiden taidehistoriallinen instituutti (RKD), jolle Iconclass-järjestelmä siirtyi Alankomaiden kuninkaalliselta tiedeakatemialta (KNAW). Järjestelmää käytetään kansainvälisesti useissa taidemuseoissa ympäri maailmaa. Suuria Iconclass-järjestelmää käyttäen luokiteltuja aineistoja ovat muun muassa saksalaisen Foto Marburg -instituutin ja amerikkalaisen Getty-säätiön kokoelmat (Anon, 2004).

Valtion taidemuseo on suomentanut Iconclass-järjestelmän suomalaisia museoita varten. Suomennos on osittainen, mutta kattaa järjestelmän keskeiset osat. Järjestelmää on toistaiseksi sovellettu kokeiluluontoisesti, mutta sen laajamittaisempaa käyttöönottoa valmistellaan.

Iconclass-järjestelmää käytetään läheisessä yhteistyössä museon kokoelmätietokantojen kanssa. Iconclass-järjestelmän avulla kokoelmaan voidaan tehdä aiheenmukaisia hakuja sekä kuvata uusien kokoelmaan lisättävien teosten sisältöä. Lüscher

(2004, s. 22) jakaakin Iconclass-järjestelmän käyttäjät käyttötarkoituksen perusteella loppukäyttäjiin, eli kuvien etsijöihin, ja luokittelijoihin. Tämän lisäksi voidaan kolmantena käyttäjäryhmänä pitää Iconclass-järjestelmää hallinnoivia ylläpitäjiä. Nämä kolme käyttäjäryhmää kuvataan lyhyesti seuraavassa:

Loppukäyttäjät on käyttäjäryhmistä heterogeenisin. Joukkoon kuuluu niin taidehistorian ammattilaisia kuin muita taiteen ystäviäkin. Loppukäyttäjän tavoitteena on löytää etsimänsä teos museon kokoelmista. Luokittelujärjestelmä on hänelle vain apuväline tämän tavoitteen saavuttamisessa. Erityisesti yleisölle suunnatuissa käyttöliittymissä luokittelujärjestelmä saattaa olla jopa kokonaan piilotettu käyttäjältä.

Luokittelijat ovat museoissa työskenteleviä taidehistorian ammattilaisia. Luokittelijan tavoitteena on kuvata teoksen sisältö mahdollisimman tarkasti, jotta sen hakeminen museon tietokannasta olisi mahdollista. Hän käyttää Iconclass-järjestelmää ensisijaisesti teoksen sisältöä kuvailevien luokittelukoodien hakemiseen ja saattaa joskus tehdä ylläpitäjille ehdotuksia uusien luokittelukoodien lisäämiseksi järjestelmään.

Ylläpitäjät vastaavat luokittelujärjestelmän hallinnasta. He arvioivat järjestelmää koskevat muutosehdotukset ja hyväksyvät luokittelujärjestelmään tehtävät lisäykset ja muutokset. Heidän tavoitteenaan on pitää järjestelmä yhtenäisenä ja ajantasaisena.

3.2 Iconclass-järjestelmän osat

Iconclass-järjestelmä koostuu seuraavista kolmesta osasta (Anon, 2004):

Hierarkkinen luokittelujärjestelmä, joka koostuu luokittelukoodista yli 28 000 taideteosten sisällönkuvailuun liittyvälle käsitteelle. Luokittelujärjestelmä on keskeinen sekä teosten haun että luokittelun kannalta, sillä se mahdollistaa haun teoksen aihepiirin ja sisällön mukaan.

Aakkosellinen indeksi sisältää asiasanalistan, jonka avulla luokittelija voi etsiä sopivia luokittelujärjestelmän koodeja kuvaillessaan teoksen sisältöä. Indeksia

voidaan hyödyntää myös haussa etsimällä tiettyä asiasanaa vastaavia luokittelukoodeja.

Bibliografia sisältää viitetietoja järjestelmää kehitettäessä käytettyihin taidehistoriallisiin lähteisiin. Ylläpitäjien lisäksi bibliografiasta on apua myös luokittelijoille, jotka voivat tarvittaessa tarkistaa käyttämänsä luokittelukoodin tarkemman merkityksen viitteiden avulla.

3.3 Luokittelujärjestelmän perusrakenne

3.3.1 Notaatiot ja määritteet

Teosten luokitteluun Iconclass-järjestelmässä käytetään tarkasti määriteltyä syntaksia noudattavia luokittelukoodeja, joita kutsutaan notaatioiksi (*notation*). Rakenteeltaan notaatiot ovat aakkosnumeerisia merkkijonoja. Kuhunkin notaatioon liittyy sanallinen määrite (*textual correlate*), joka kertoo notaation merkityksen. Kuva 3.1 esittää esimerkin luokittelukoodista. (Anon, 2004; Anttonen, 2003, s. 1)

$\underbrace{25G21 (APPLE)}_{\text{notaatio}}$	$\underbrace{\text{hedelmät: omena}}_{\text{määrite}}$
--	--

Kuva 3.1: Luokittelukoodin rakenne Iconclass-järjestelmässä

Alkuperäisessä Iconclass-järjestelmässä määritteet ovat englanniksi, mutta järjestelmästä on käytöstä erikielisiä käännöksiä. Koska varsinainen luokittelu tehdään notaatioiden perusteella, on järjestelmä tältä osin kieliriippumaton. Luokittelujärjestelmästä on käytössä myös Valtion taidemuseossa tehty suomennos, joka sisältää käännökset n. 4 000 keskeiselle määritteelle. (Lüscher, 2004, s. 18)

3.3.2 Hierarkkinen rakenne

Luokittelujärjestelmä noudattaa systemaattista menettelyä luokkahierarkian muodostamisessa. Huipulla oleva kaikkeuden käsite jaetaan Iconclass-järjestelmässä ensin kymmeneen pääluokkaan. (Hyvönen, 2005b, s. 11) Näistä luokat 1–5 pyrkivät kattamaan kaiken yleisesti kuvattavissa olevan ja luokat 6–9 keskittyvät kerronnal-

lisiin aihepiireihin, erityisesti Raamattuun ja klassiseen mytologiaan. Lisäksi järjestelmään on myöhemmin lisätty abstraktin taiteen kuvailuun tarkoitettu luokka 0. (Anon, 2004) Pääluokat suomennettuine määritteineen on esitetty taulukossa 3.1.

Taulukko 3.1: Iconclass-luokittelujärjestelmän kymmenen pääluokkaa (Anttonen, 2003)

0 abstrakti taide, ei-esittävä taide, non-figuratiivinen taide	5 abstraktit ideat ja käsitteet
1 uskonto ja uskomusmaailma	6 historia
2 luonto	7 Raamattu
3 ihminen	8 kirjallisuus
4 yhteiskunta, sivilisaatio, kulttuuri	9 klassinen mytologia ja antiikin historia

Iconclass-järjestelmän notaatioiden ensimmäisenä merkinä on aina jonkin pääluokan 0–9 numerokoodi. Kukin pääluokka jakaantuu edelleen systemaattista rakennetta noudattaen pääluokan merkitystä täsmentäviin alaluokkiin, joita esitetään myös numerokodeilla (Anon, 2004). Esimerkiksi pääluokan 3 alaluokat on esitetty taulukossa 3.2.

Taulukko 3.2: Pääluokan 3 alaluokat Iconclass-luokittelujärjestelmässä

31 ihminen biologisena olentona	34 ihminen ja eläin
32 ihmistyytit; kansat ja kansallisuudet	35 pastoraalit ja idyllit, Arkadia-maisemat
33 ihmisten väliset suhteet	36 ihminen ja ympäristö

Hierarkian kahden ylimmän tason osalta luokittelujärjestelmä muistuttaa rakenteeltaan läheisesti yleisten kirjastojen luokitusjärjestelmää (Anon, 2003). Nämä yhtäläisyydet kuitenkin loppuvat siirryttäessä syvemmälle Iconclass-järjestelmän hierarkiassa. Hierarkian kolmas taso esitetään lisäämällä notaatioon kirjain väliltä A–Z luukuunottamatta kirjainta J, joka on jätetty pois luettavuuden parantamiseksi. Näin saavutetaan suurempi rinnakkaisten vaihtoehtojen määrä kuin numeroa käytettäessä (Anon, 2004).

Luokittelujärjestelmän syvempien tasojen notaatiot muodostetaan lisäämällä notaation perään numeroita (Anon, 2004). Samalla notaation merkitys tarkentuu hy-

vinkin yksityiskohtaisiin asioihin ja tapahtumiin (Lüscher, 2004, s. 17). Esimerkiksi notaatio 41A163 tarkoittaa omakotitaloa tai mökkiä ja notaation 73B14 määritteenä on ”enkelin ilmestyminen paimenille yöllä; ilmoitus Kristuksen syntymästä, joukko enkeleitä laulaa taivaalla”.

3.3.3 Semanttiset suhteet

Koska Iconclass-järjestelmän notaatiot rakentuvat hierarkkisesti, sisältää jokainen notaatio aina implisiittisesti myös kaikkien ikonografisesti laajempien yläluokkien notaatiot (Anon, 2005b). Luokittelujärjestelmän runko muodostaa näin puurakenteen, jossa luokkien välillä on hierarkkiset LT- ja ST-suhteet.

Hierarkkiset suhteet ovat tyypiltään mm. osa–kokonaisuus- tai ala–yläkäsitesuhteita (Hyvönen, 2005b, s. 4). Iconclass-järjestelmässä ei kuitenkaan erotella näitä suhdetyyppejä toisistaan (Wielinga et al., 2001). Esimerkiksi viikontähtiä kokonaisuutena merkitsevän notaation 23Q alta löytyvät notaatiot yksittäisille viikontähtiville, kun taas notaation 34B233 ”vesilinnut kotieläiminä” alakäsitteinä on eri lintulajeja: joutsenet, hanhet ja ankat. Pelkän Iconclass-syntaksiin sisältyvän informaation perusteella ei näin ollen voida ottaa kantaa hierarkkisten suhteiden laatuun, vaan näiden selvittäminen edellyttää notaatioihin liittyvien sanallisten määritteiden tarkastusta.

Hierarkkisten suhteiden lisäksi Iconclass-järjestelmän luokkien välillä on systemaattiseksi viitteiksi kutsuttuja ristiviitteitä, jotka liittävät yhteen toisiinsa läheisesti liittyviä luokkia luokittelujärjestelmän eri osista. Viitettä ei kaikissa tapauksissa ole luokittelujärjestelmässä liitetty molempien luokkien yhteyteen, mutta vaikuttaa siltä, että viittaukset ovat luonteeltaan kaksisuuntaisia assosiatiivisia RT-suhteita.

Luokittelujärjestelmästä löytyy myös KÄYTÄ/KORVAA-suhteita. Näille suhteille ei kuitenkaan ole Iconclass-järjestelmässä mitään formaalia rakennetta, vaan ne on ilmaistu ainoastaan notaatioihin liittyviin sanallisiin määritteisiin kirjattuina huomautuksina. Esimerkiksi notaation 85A määrite ilmaisee KÄYTÄ-suhteen: ”Aisopoksen faabelit : Käytä 85(...)”. Käänteistä KORVAA-suhdetta ei ole eksplisiittisesti ilmaistu.

3.4 Luokittelujärjestelmän laajennukset

Notaatioista ja niitä kuvaavista määritteistä muodostuvan luokittelujärjestelmän hierarkkisen rungon lisäksi Iconclass-järjestelmään kuuluu myös joukko laajennuksia. Nämä mahdollistavat täsmällisempien merkitysten ilmaisemisen notaatioiden avulla. Laajennusten tarjoamien tarkempien ja ilmaisuvoimaisempien notaatioiden hintana on kuitenkin järjestelmän rakenteen monimutkaistuminen. (Anon, 2004)

Seuraavissa luvuissa esitellään luokittelujärjestelmän neljä laajennusta, joista kukin tuo omat lisäyksensä notaatioissa käytettävään syntaksiin.

3.4.1 Nimi sulkeissa

Tietyissä luokittelujärjestelmän osissa esiintyy mahdollisuus tarkentaa notaatiota liittämällä siihen nimi sulkeissa (*bracketed text*). Näin menetellään useiden muun muassa historiallisia henkilöitä, pyhimyksiä sekä kasvi- ja eläinlajeja tarkoittavien notaatioiden kohdalla. Mahdollisuus laajennuksen käyttöön ilmaistaan notaatiolla, jonka lopussa on sulkeissa kolme pistettä. Myös notaation määrite kertoo laajennuksen käyttömahdollisuudesta. (Anon, 2004; Anttonen, 2003) Esimerkiksi notaation 25F23(. . .) määritteenä on ”petoeläimet (NIMEN kanssa)”.

Useimmiten sulkeisiin suuraakkosin kirjoitettava nimi valitaan erityisestä kontrolloidusta sanastosta. Sanastossa esiintyvät nimet ovat aina englanninkielisiä, jotta notaatiot säilyisivät yksikäsitteisinä. Myös vapaan asiasanan käyttö on sallittua, mikäli sanastossa ei ole sopivaa asiasanaa. Vapaa asiasana ei kuitenkaan varsinaisesti ole osa luokitettujärjestelmää, vaan järjestelmän kannalta notaatio rinnastuu ylemmän ilman sulkeita olevaan notaatioon. (Anon, 2004; Lüscher, 2003, s. 29) Esimerkiksi taulukossa 3.3 esitettyä petoeläinten sanastoa käyttämällä voidaan tiikerille muodostaa notaatio 25F23(TIGER). Paikannimi Helsinki voidaan merkitä vapaata asiasanaa käyttäen notaatiolla 61E(HELSINKI) ”kaupunkien ja kylien nimet (NIMEN kanssa)”, sillä paikannimille ei järjestelmässä ole omaa sanastoa. Tämä notaatio on kuitenkin epätarkka, koska Suomessa on kaksi Helsinki-nimistä paikkaa. (Salminen, 2006, s. 10)

Taulukko 3.3: Nimi sulkeissa -laajennuksessa käytettävä petoeläin-sanasto. Suomennos osittainen

25F23(BADGER)	beasts of prey, predatory animals: badger
25F23(BEAR)	petoeläimet: karhu
25F23(CHEETAH)	beasts of prey, predatory animals: cheetah
25F23(ERMINE)	beasts of prey, predatory animals: ermine
25F23(FERRET)	beasts of prey, predatory animals: ferret
25F23(FOX)	petoeläimet: kettu
25F23(HYENA)	petoeläimet: hyeena
25F23(ICHNEUMON)	beasts of prey, predatory animals: ichneumon
25F23(LEOPARD)	petoeläimet: leopardi
25F23(LION)	petoeläimet: leijona
25F23(LYNX)	petoeläimet: ilves
25F23(MANGOUSTE)	beasts of prey, predatory animals: mangouste
25F23(MARTEN)	beasts of prey, predatory animals: marten
25F23(OTTER)	beasts of prey, predatory animals: otter
25F23(PANTHER)	petoeläimet: pantteri
25F23(SABLE)	beasts of prey, predatory animals: sable
25F23(TIGER)	petoeläimet: tiikeri
25F23(WEASEL)	beasts of prey, predatory animals: weasel
25F23(WOLF)	petoeläimet: susi

Iconclass-järjestelmän suomennoksessa käytännöt tämän laajennuksen osalta eivät ole vielä vakiintuneet. Nykyinen suomennos ei kata kaikkia kontrolloituihin sanastoihin sisältyviä määritteitä, joten osalle notaatioista on olemassa ainoastaan englanninkieliset vastineet. Myöskään käytäntö vapaiden asiasanojen käytöstä notaatioissa ei ole vielä vakiintunut. Suomennostyöryhmä kuitenkin on päättänyt suositella suomenkielisten asiasanojen käyttöä notaatioissa silloin kuin kontrolloitua sanastoa ei ole käytettävissä. (Lüscher, 2003, s. 29)

Anon (2004) kuvaa nimi sulkeissa -laajennuksen tavaksi irrottautua luokittelujärjestelmän hierarkkisesta rakenteesta. Laajennus lisääkin rinnakkaisten vaihtoehtojen määrää, ja täten vapauttaa osin notaatioiden normaalin syntaksin rajoitteista. Hierarkkiset LT- ja ST-suhteet laajennusta käyttävän notaation ja luokittelujärjestelmän

rungon välillä säilyvät kuitenkin useimmissa tapauksissa hyvin selkeinä, kun nimi sulkeissa -laajennusta käytetään esimerkiksi kasvi- ja eläinlajien yksilöintiin.

3.4.2 Kaksoiskirjaimet

Kaksoiskirjaimia (*doubling of the letter*) käytetään vastakohtien ilmaisemiseen Iconclass-järjestelmässä. Tyypillinen esimerkki tästä on sukupuolen ilmaiseminen notaatiossa: notaatio 61B merkitsee miespuolisia historiallisia henkilöitä ja notaatio 61BB vastaavasti naispuolisia. Notaation käyttö rajoittuu kuitenkin vain tiettyihin luokittelujärjestelmän osiin, ja esimerkiksi sukupuolen erotteluun voidaan käyttää myös tavanomaisia notaatioita, kuten notaatioissa 96C ja 96D, jotka merkitsevät mies- ja naispuolisia roomalaisen taruston hahmoja.

Kaksoiskirjainten käytöllä voidaan määrittää tietyissä päähierarkian osissa kaksi rinnakkaista alipuuta, joihin kuuluvien notaatioiden merkitys on keskenään vastakohtainen tai käänteinen (Anon, 2004). Aitchison et al. (2000, ref. Hyvönen, 2005b, s. 12) mainitsee vastakohtaisuuden yhtenä assosiatiivisen suhteen tyyppinä. Tätä rinnakkaisten notaatioiden välistä RT-suhdetta ei kuitenkaan ole eksplisiittisesti ilmaistu luokittelujärjestelmässä.

3.4.3 Rakenteellinen numerointi

Erityisesti luokittelujärjestelmän luokissa 11 "kristinusko", 12 "ei-kristilliset uskonnot" ja 9 "klassinen mytologia ja antiikin historia" on runsaasti erilaisiin ryhmiin kuuluvia henkilöahmoja ja heidän toimintaansa kuvaavia luokkia. Rakenteellisen numeroinnin (*structural digits*) tarkoituksena on määrittää yhtenäinen käytäntö notaatioiden muodostamiselle samaan ryhmään kuuluville hahmoille. Esimerkiksi klassisen historian henkilöihin liittyvissä notaatioissa voidaan notaation loppuun liittää numero taulukossa 3.4 esitetyltä listalta. Notaatio 98B(ALEXANDER THE GREAT)4 merkitsee näin Aleksanteri Suuren aggressiivisia suhteita ja vastaavasti Julius Caesarin aggressiivisille suhteille saadaan notaatio 98B(CAESAR)4.

Pelkän yksittäisen notaation perusteella ei voida päätellä, onko siinä käytetty rakenteellista numerointia vai ei, sillä laajennusta käyttävät notaatiot eivät eroa muo-

Taulukko 3.4: Rakenteellinen numerointi klassisen historian henkilöiden notaatioissa

98B(...)1	klassiseen historiaan liittyvän henkilön varhaislapsuus ja -nuoruus sekä kasvatus
98B(...)2	klassiseen historiaan liittyvän henkilön rakkausjutut tai lemmenseikkailut
98B(...)3	esitykset, joihin klassiseen historiaan liittyvän henkilön NIMI voidaan yhdistää
98B(...)4	klassiseen historiaan liittyvän henkilön aggressiiviset suhteet
98B(...)5	klassiseen historiaan liittyvän henkilön ei-aggressiiviset toimet
98B(...)6	klassiseen historiaan liittyvän henkilön kärsimys, epäonni, onnettomuus
98B(...)7	kuolemanjälkeiset tapahtumat ~ klassiseen historiaan liittyvä henkilö
98B(...)8	tarunomaiset, selkeästi legendanomaiset lisäpiirteet ~ klassiseen historiaan liittyvä henkilö
98B(...)9	klassiseen historiaan liittyvän henkilön muotokuva

doltaan mitenkään luokittelujärjestelmän perusrakenteeseen kuuluvista notaatioista. Laajennuksen käyttö selviää vain seuraamalla notaation määrittämää hierarkkista polkua luokittelujärjestelmässä.

Koska rakenteellinen numerointi määrittää rinnakkaisiin alipuihin kuuluville notaatioille keskenään samankaltaisen merkityksen, mahdollistaa se samalla poikkileikkausten tekemisen läpi luokittelujärjestelmän hierarkian (Anon, 2004). Nämä poikkileikkaukset ovat luonteeltaan assosiatiivisia RT-suhteita, mutta ne esiintyvät luokittelujärjestelmässä vain implisiittisesti.

3.4.4 Avaimet

Avaimet (*keys*) ovat luokittelujärjestelmän määrättyihin osiin liitetty menetelmä tämentää ja muuttaa notaation merkitystä. Avain on notaation perään sulkeisiin liitettävä +-merkillä alkava numerokoodi. Avaimen merkityksen tulkitsemiseksi on tiedettävä mihin notaatioon se liittyy, sillä samalla avaimella voi olla eri merkityksiä riippuen notaatiosta, johon se on liitetty. Avain voi esimerkiksi tarkentaa notaatiota,

kuten maisemaa merkitsevään notaatioon 25H liitetty avain (+1), joka kertoo, että maisemassa esiintyy ihmis- tai eläinhahmoja. Myös tapahtumaa voidaan kuvata avaimen avulla. Notaation 34B12 merkitys on kissa. Kun siihen liitetään avain (+1), joka tässä yhteydessä tarkoittaa eläimen ruokkimista, saadaan kissan ruokkiminen -tapahtumalle notaatio 34B12(+1). Kolmas käyttötapa avaimille on prosessien kuvaaminen. Viljaa tarkoittavaan notaatioon 47I111 liitettynä avain (+82) merkitsee kylvöprosessia.

Rekursiiviset avaimet mahdollistavat avaimen merkityksen tarkentamisen liittämällä avaimeen numeroita muualla luokittelujärjestelmässä määritellyistä avaimista. Liitettyjä numeroita kutsutaan jonoksi (*queue of key*). Menettelyn tarkoituksena on vähentää tarpeetonta toisteisuutta avainten määrittelyissä. (Anon, 2005a) Esimerkiksi notaation 25F "eläimet" yhteydessä on määritelty avain (+641), joka merkitsee eläimen teurastamista. Karjaeläimiä ei luokittelujärjestelmässä ole kuitenkaan sijoitettu muiden eläinten yhteyteen, vaan osaksi maatalouteen liittyviä notaatioita. Näin ollen sikaa merkitsevään notaatioon 47I212 ei voida suoraan liittää teurastamista merkitsevää avainta, koska 47-alkuisille notaatioille on käytettävissä taulukossa 3.5 esitettävät avaimet. Ratkaisun tarjoaa rekursiivinen avain (+9), joka antaa mahdollisuuden viitata notaation 25F yhteydessä määriteltyihin avaimiin. Näin sian teurastamiselle saadaan notaatio 47I212(+9641).

Taulukko 3.5: Käsityö ja teollisuus -luokkaan liitettävissä olevat avaimet. Suomennos osittainen

47(+0)	variant
47(+1)	paikka ~ käsityö ja teollisuus
47(+2)	abourers ~ crafts, industries, agriculture
47(+3)	technical aspects ~ crafts, industries, agriculture
47(+4)	materials processed ~ crafts, industries, agriculture
47(+5)	basic industrial processes
47(+6)	before and after production
47(+7)	forms of production and organization
47(+8)	agricultural processes
47(+9)	zoological aspects of domestic cattle and poultry

Pelkän avaimen numerokoodin perusteella ei voida päätellä, onko avainta laajennettu rekursiivisen avaimen ja sen tarjoaman jonon avulla. Tämä voidaan selvittää vain avaimen määrittämää hierarkkista polkua seuraamalla. Tässä suhteessa laajennus on siis samankaltainen kuin notaatioiden laajentamiseen käytettävä rakenteellinen numerointi.

Avaimet ovat luokittelujärjestelmän laajennuksista laajin ja samalla semanttiselta rakenteeltaan monimutkaisin. Avain liittyy aina johonkin luokittelujärjestelmän päähierarkian alipuuhun (Anon, 2004). Avain voi esiintyä vain tähän joukkoon kuuluvien notaatioiden yhteydessä, missä sen merkitys on määritelty.

Sisäisesti myös avaimet muodostavat vastaavaan tapaan numerokooodeista rakentuvia avaimen merkitystä täsmentäviä hierarkioita kuin luokittelujärjestelmän päähierarkian notaatiotkin. Samaan hierarkiaan kuuluvien avainten välillä on näin myös LT- ja ST-suhteet.

3.5 Aakkosellinen indeksi

Aakkosellisen indeksin englanninkielinen versio sisältää 14 000 asiasanaa, joihin liittyy viittauksia luokittelujärjestelmän notaatioihin. Suomennetussa versiossa asiasanoja on noin 11 900. Indeksillä on myös sisäistä semanttista rakennetta, sillä se sisältää toisiinsa läheisesti liittyvien asiasanojen välisiä ristiviitteitä.

Notaatioon liittyvä asiasanoitus ei kuitenkaan välttämättä kuvaa täydellisesti notaation merkitystä. Asiasanoitus saattaa olla puutteellinen ja sisältää vain osan notaatioon liittyvistä elementeistä. Esimerkiksi notaatioon 47D ”koneet, koneen osat, työkalut ja laitteet” liitetystä asiasanoista ”yhteiskunta, sivilisaatio, kulttuuri, käsityö, teollisuus, teollisuustoiminta, ammatti, työ, tehtävä” yksikään ei suoraan viittaa koneisiin ja laitteisiin.

Tapahtumia ja toimintaa esittävissä notaatioissa asiasanoitus ei kerro, millainen rooli eri asiasanoilla on suhteessa tapahtumaan. Esimerkiksi notaatio 34A3 merkitsee

koiran kouluttamista ja siihen on liitetty asiasanat ”ihminen, eläin, ammatti, työ, tehtävä, koira, koulutus, valmennus”. Ihminen pystyy päättämään tästä sanalis-
tasta, että kyseessä on tilanne, jossa ihminen toimii valmentajana ja kouluttaa koi-
raa. Koneelle sen sijaan myös vaihtoehto, jossa koira kouluttaa ihmistä olisi täysin
yhtä todennäköinen, mikäli se ei kykene tulkitsemaan tilanteeseen liittyviä semant-
tisia rooleja.

Aakkosellisen indeksin suomennos on tehty kääntämällä englanninkielisen version
asiasanalista suomeksi. Näin asiasanat kytkeytyvät luokittelujärjestelmän notaatioi-
hin englanninkielisten vastineidensa kautta. Koska kaikille asiasanoille ei ole tark-
kaa suomennosta tai sanalle on useita käännösvaihtoehtoja, saattaa suomenkielinen
indeksi tarjota joissain tapauksissa notaatioita, joiden yhteys asiasanaan ei ole ko-
vin ilmeinen. Esimerkiksi haku asiasanalla kylpytakki antaa tulokseksi myös kaikki
viittoihin ja kaapuihin liittyvät notaatiot, koska englanninkielisen indeksin asiasana
robe on suomennokseksi on annettu nämä kaikki kolme merkitystä. Sen sijaan vas-
taavan englanninkielisessä indeksissä asiasanalla *bath-robe* tehdyn haun tuloksena
on ainoastaan kylpytakkia merkitsevä notaatio 31A5464.

3.6 Semanttiset sovellukset

Iconclass-järjestelmää on hyödynnetty semanttisissa sovelluksissa aiemminkin. Nä-
mä sovellukset eroavat tästä diplomityöstä siinä, että niissä ei ole pyritty kuva-
maan Iconclass-järjestelmän koko semanttista rakennetta vaan Iconclass-järjestel-
mästä on hyödynnetty vain osaa laajemman kokonaisuuden yhteydessä. Seuraa-
vassa esitellään kaksi Iconclass-järjestelmää käyttävää semanttista sovellusta.

Wielemaker et al. (2003) esittää AAT-, WordNet- ja Iconclass-järjestelmiin perustu-
van sovelluksen taideteosten annotointiin ja hakuun. Iconclass-luokittelujärjestel-
män hierarkkista käsitteistöä käytetään osana teosten sisällönkuvailua. Sovelluk-
sessa hyödynnetään myös ontologioiden välisiä käsitevastaavuuksia esimerkiksi
Iconclass- ja AAT-järjestelmien välillä. Ontologiat on kuvattu RDF Schemaa käyt-
täen.

Monta kertaa kaupunkiin -projektissa muodostettiin ”kaupunkiontologia” yhdistä-

mällä Iconclass-notaatioihin liittyvät määritteet aihekartan avulla toisiinsa (Heikka, 2006). Näin syntyvä ontologia korostaa assosiatiivisia suhteita, jotka eivät muuten ole pääosin hierarkkiseen rakenteeseen perustuvan Iconclass-järjestelmän vahvin puoli.

Tällä hetkellä Iconclass-järjestelmän semanttisiin sovelluksiin liittyvää tutkimusta on FinnONTO-projektin lisäksi meneillään ainakin Alankomaissa. van Gendt et al. (2006) esittelee sovelluksen, jossa Iconclass-järjestelmä on liitetty toiseen kuvataiteen luokittelujärjestelmään, kun taas KulttuuriSammossa tavoitteena on yhdistää erityyppisiä kulttuurisisältöjä toisiinsa.

3.7 Yhteenveto

Kuvataideteosten aiheenmukaiseen luokitteluun käytettävä Iconclass-järjestelmä koostuu hierarkkisesta kymmeneen pääluokkaan jakaantuvasta luokittelujärjestelmästä, aakkosellisesta indeksiosasta ja taidehistoriallisesta bibliografiasta. Luokittelujärjestelmä käyttää teosten luokitteluun aakkosnumeerisia notaatioita, joihin liittyy merkitystä kuvaava sanallinen määrite.

Iconclass-luokittelujärjestelmän semanttiset suhteet ovat luonteeltaan samankaltaisia kuin tyypillisissä asiasanastoissa. Luokittelujärjestelmän rungon muodostavat hierarkkiset LT- ja ST-suhteet, joiden lisäksi samankaltaisten luokkien välillä esiintyy assosiatiivisia RT-suhteita.

Nimi sulkeissa, kaksoiskirjaimet ja rakenteellinen numerointi -laajennukset noudattavat luokittelujärjestelmän rungon hierarkkista rakennetta. Ne määrittävät uusia laajennuksen sisäisiä assosiatiivisia suhteita. Avaimet-laajennus määrittelee oman päähierarkiasta erillisen hierarkkisen rakenteensa, jossa määritellyt avaimet täsmenävät päähierarkian tiettyyn alipuuhun kuuluvien notaatioiden merkitystä.

4 HAASTEET KONESEMANTIICALLE

Tässä luvussa käsitellään Iconclass-luokittelujärjestelmän muuntamista konesemanttisesti käsiteltävään muotoon. Ensiksi tarkastellaan Iconclass-luokittelujärjestelmän semanttista rakennetta ja sen esittämistä SKOS Core -mallia käyttäen. Tämän jälkeen käsitellään yksilöivien tunnisteiden muodostamista ontologian käsitteille. Lopuksi tutkitaan vaihtoehtoja ICONCLASS-luokittelun merkityksen esittämistä muiden ontologioiden avulla.

4.1 Luokittelujärjestelmän rakenteen kuvaaminen

Tämän diplomityön työhypoteesina oli, että SKOS Core -mallia voidaan soveltaa Iconclass-luokittelujärjestelmän esittämiseen ontologiana. Hypoteesin perusteena oli mallin keskittyminen erilaisten käsittemallien, kuten tesaurusten ja luokittelujärjestelmien perusrakenteen kuvaamiseen (Brickley & Guha, 2004; Miles & Brickley, 2005).

SKOS Core -malli pyrkii esittämään rakenteita hyvin yleisellä tasolla. Se keskittyy käsitteiden ja niiden välisten suhteiden kuvaamiseen. Mallia on kehitetty pääasiassa tesaurusten mallintamisen tarpeista lähtien, mistä johtuen käsitteitä kuvataan ensisijaisesti luonnollisella kielellä esitetyillä termeillä. Koska Iconclass-järjestelmä on lähtökohdaltaan erilainen ja siinä on pyritty irti luonnollisen kielen luokittelulle asettamista rajoitteista käyttämällä tunnisteena ensisijaisesti kieliriippumatonta notaatiota ja sanallista määritettä vain kuvailevassa roolissa, on SKOS Core -mallia Iconclass-järjestelmään sovellettaessa ratkaistava, miten luokittelujärjestelmän rakenteet voidaan mallintaa SKOS Core -mallin tarjoamia luokkia ja ominaisuuksia käyttäen.

4.1.1 Käsitteiden nimikkeet

SKOS Core -mallissa käsitteen ensisijaisena nimikkeenä käytetään `skos:prefLabel` ominaisuutta, jonka arvo määritelmän mukaan sisältää resurssin suositeltavan sanallisen tunnisteen luonnollisella kielellä. Sama vaatimus kohteen ihmisluettavasta nimestä löytyy myös `rdfs:label` ominaisuudesta, jonka aliominaisuus

`skos:prefLabel` on. Iconclass-järjestelmä käyttää luokan ensisijaisena tunnisteena kieliriippumatonta notaatiota, joka ei yksinään ole ihmisen luettavaksi soveltuva tunniste.

Erään tavan notaatioiden merkintäongelman ratkaisemiseksi tarjoaa `skos:prefLabel` ominaisuuden arvon kielen merkitään käytettävä `xml:lang`-attribuutti (Bray et al., 2006). Tämän attribuutin arvona voidaan käyttää ISO 639 -standardissa määriteltyjä kielikoodia (Alvestrand, 2001). Vuoden 2006 alussa tähän standardiin lisättiin uusi kielikoodi `zxx`, jolla voidaan ilmaista ettei kyse ole kielellisestä sisällöstä (Anon, 2006a). Luokittelujärjestelmien julkaisemista SKOS Core -mallin mukaisessa muodossa käsittelevän ohjeen luonnoksessa suositellaankin `zxx`-kielikoodilla merkityn `skos:prefLabel` ominaisuuden käyttämistä, mikäli luokittelujärjestelmässä käytetään notaatioita luokkien tunnisteina (Voss, 2006). Koska kyse on alustavasta luonnoksesta, käytetty merkintätapa voi vielä muuttua.

Parhaiten notaation esittämiseen SKOS Core -mallissa soveltuisi ID-tyyppinen merkijonotunniste. Tällä hetkellä malli ei tarjoa tällaista ominaisuutta, mutta sen tarpeellisuudesta on keskusteltu mallin kehittäjien keskuudessa, koska useat luokittelujärjestelmät noudattavat samankaltaista merkintätapaa (Svensson, 2006).

Vaihtoehtona `skos:prefLabel` ominaisuuden arvoksi olisi lähinnä notaatioon liittyvä sanallinen määrite. Useimmat määritteet eivät kuitenkaan ole tiiviitä yksiselitteisiä otsikoita luokalle, vaan pitempiä luokan käyttöä kuvailevia tekstejä. Tästä syystä määrite soveltuukin paremmin käsitteen sisältöä kuvaavan `skos:definition` ominaisuuden arvoksi.

SKOS Core -määrittely edellyttää, että kaikilla käsitteillä on täsmälleen yksi `skos:prefLabel` ominaisuus kieltä kohden. Sekä notaatio että määrite voitaisiin siis esittää `skos:prefLabel` ominaisuuden arvoina, mikäli niillä on eri kielikoodit. Tämä ratkaisu saattaisi kuitenkin vaikeuttaa notaatioiden esittämistä ontologiaa käyttävissä sovelluksissa, sillä se edellyttäisi saman ominaisuuden kahden erikielisen arvon esittämistä samanaikaisesti.

Tässä työssä notaatioiden esittämiseen päädyttiin käyttämään `zxx`-kielikoodilla mer-

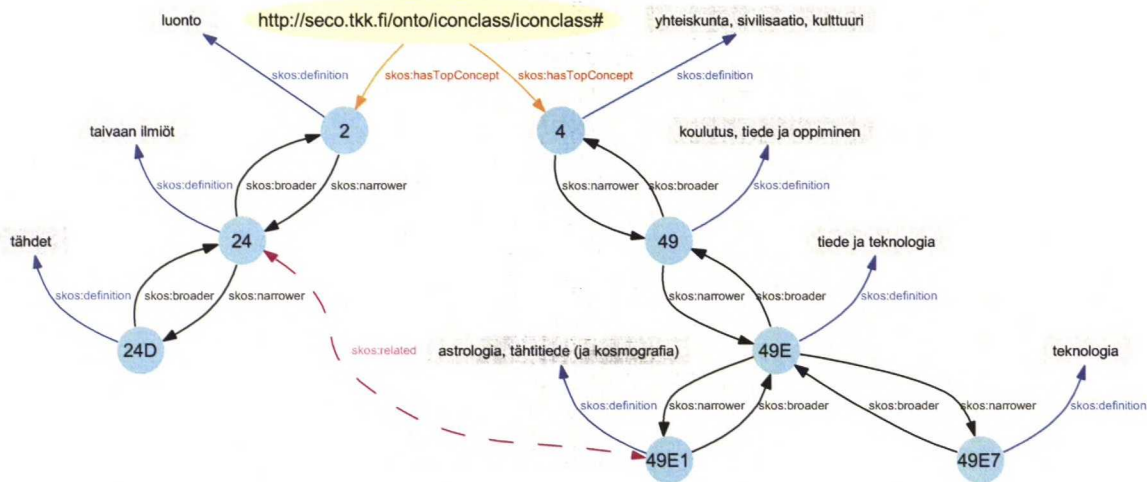
kittyä skos:prefLabel ominaisuutta ja määritteiden esittämiseen skos:definition ominaisuutta. Nämä SKOS Core -mallin ja Iconclass-luokittelujärjestelmän vastaavuudet käsitteiden kuvaamisessa on esitetty taulukossa 4.1.

Taulukko 4.1: Käsitteen nimikkeen ja kuvauksen esittäminen Iconclass-luokittelujärjestelmässä ja SKOS Core -mallissa

	Iconclass	SKOS Core
<i>ensisijainen nimike</i>	notaatio	skos:prefLabel
<i>sanallinen kuvaus</i>	määrite	skos:definition

4.1.2 Semanttiset suhteet

Iconclass-luokittelujärjestelmän hierarkkinen runko sisältää LT- ja ST-suhteita, joten se voidaan kuvata SKOS Core -mallissa skos:broader ja skos:narrower ominaisuuksia käyttäen. Luokkien väliset ristiviittaukset ovat RT-suhteita, jotka voidaan esittää skos:related ominaisuuden avulla. Kuva 4.1 esittää osan Iconclass-luokittelujärjestelmän hierarkiasta kuvattuna SKOS Core -mallin mukaisesti.



Kuva 4.1: Osa Iconclass-luokittelujärjestelmän pääluokista 2 ja 4 kuvattuna SKOS Core -mallin mukaisesti

Iconclass-järjestelmässä RT-suhde voi esiintyä minkä tahansa kahden luokan välillä mukaan lukien luokat, joiden välillä on myös hierarkkinen ST-suhde. Tapauksia, joissa luokkien välillä on molemmat suhteet on luokittelujärjestelmässä vain muutama. Näissä tapauksissa RT-suhteella pyritään suosittelemaan valittua yläluokkaa

tarkempaa notaatiota. Esimerkiksi luokkaan 31A ”ihmisfiguuri” on RT-suhteella liitetty 31A7 ”sukupuolet” osoittamaan mahdollisuutta tarkentaa, kumpaa sukupuolta kuvaava ihmisfiguuri on kyseessä. SKOS Core -malli ei tällä hetkellä ota kantaa tapaukseen, jossa käsitteiden välillä on sekä hierarkkinen että assosiatiivinen suhde (Miles & Brickley, 2005). Luonnos SKOS Core -mallin mukaisen aineiston yhtenäisyytsteiksi esittää kuitenkin, että käsitteiden välillä ei saisi olla sekä hierarkkista että assosiatiivista suhdetta (Miles, 2005a). Mikäli tämä rajoite otetaan mukaan SKOS Core -malliin, edellyttää Iconclass-luokittelujärjestelmän semantiikan kuvaaminen mallin laajentamista uudella assosiatiivisella suhdetyypillä, jolla ei ole rajoitetta muiden samaan käsitteeseen liittyvien suhteiden osalta.

KÄYTÄ/KORVAA-suhteet on Iconclass-järjestelmässä esitetty sanallisiin määritteisiin liitettyinä huomautuksina. SKOS Core -mallissa nämä on esitetty saman käsitteen ensisijaisena ja vaihtoehtoisena nimikkeenä skos:prefLabel ja skos:altLabel ominaisuuksien avulla. Lisäksi skos:editorialNote ominaisuuden avulla malliin voidaan liittää ihmiskäyttäjälle tarkoitettu huomautus käsitteen käytöstä. Tämä mahdollistaa notaation merkitystä kuvaavan informaation erottamisen luokituksen käyttöä koskevasta ohjeesta. Taulukko 4.2 esittää yhteenvedon luokittelujärjestelmässä esiintyvistä semanttisista suhteista ja niiden vastineista SKOS Core -mallissa.

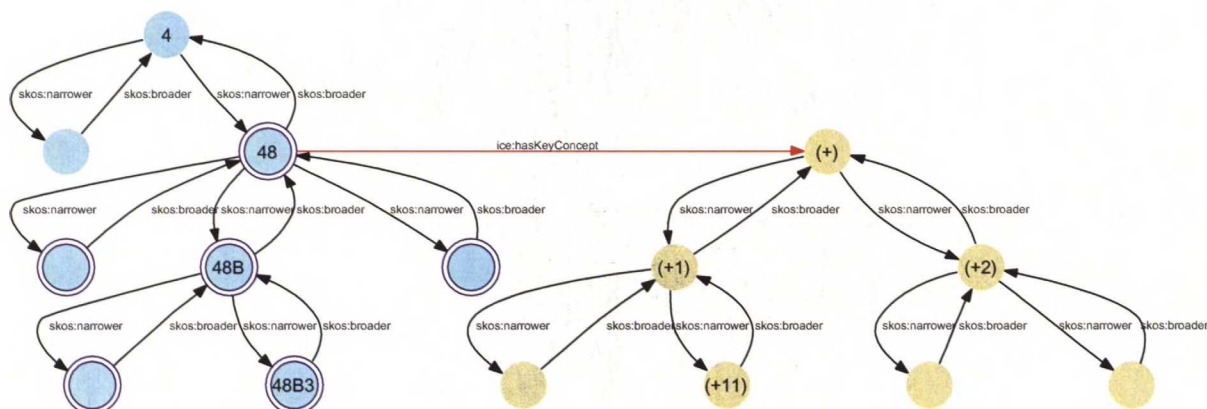
Taulukko 4.2: Semanttisten suhteiden tyypit Iconclass-luokittelujärjestelmässä ja SKOS Core -mallissa

Suhdetyyppi	Iconclass	SKOS Core
LT	notaatioiden luokkahierarkia	skos:broader
ST	notaatioiden luokkahierarkia	skos:narrower
RT	ristiviitteet notaatioiden välillä	skos:related
KÄYTÄ/KORVAA	sanallinen huomautus määritteessä	skos:prefLabel skos:altLabel skos:editorialNote

4.1.3 Avain-laajennuksen rakenne

SKOS Core -mallissa kaikki käsitteet ovat `skos:Concept`-luokan instansseja, joiden välillä voi esiintyä hierarkkisia ja assosiatiivisia suhteita. Luokittelujärjestelmän avain-laajennusta käytettäessä Iconclass-järjestelmä sisältää kahta eri tyyppiä olevia käsitteitä notaatioita ja avaimia. Sekä notaatiot että avaimet muodostavat molemmat oman hierarkkisen rakenteensa, eikä eri tyyppisiä käsitteitä voi vapaasti yhdistää toisiinsa. Avain-laajennuksen kuvaaminen edellyttääkin mahdollisuutta kahden erityyppisen käsitteen esittämiseen, mitä SKOS Core -malli ei suoraan tarjoa. Ongelma voidaan ratkaista laajentamalla mallia ja perimällä `skos:Concept`-luokasta omat luokat notaatioille ja avaimille.

Avain-laajennuksessa määriteltyt avaimet liittyvät aina tiettyyn luokittelujärjestelmän rungon haaraan, johon kuuluvien notaatioiden yhteydessä avaimia voidaan käyttää. Tämä avaimen voimassaoloalueen määrittävä luokittelujärjestelmän rungon ja avain-laajennuksen välinen suhde ei ole kuitenkaan tyypiltään tavallinen kahden käsitteen välinen hierarkkinen tai assosiatiivinen suhde, sillä avain-laajennusta voidaan käyttää myös samassa haarassa alempana olevien käsitteiden yhteydessä. Kuva 4.2 esittää esimerkin luokittelujärjestelmän rungon ja avain-laajennuksen välisestä suhteesta.



Kuva 4.2: Avain-laajennuksen suhde Iconclass-luokittelujärjestelmän runkoon. Avainta voidaan käyttää ympyröityjen käsitteiden yhteydessä

4.2 Yksilöivän tunnisteiden muodostaminen ontologian käsitteille

WWW:n arkkitehtuurissa resurssiin voidaan viitata määrittämällä sille yksilöllinen URI-tunniste (Berners-Lee et al., 2004). Tätä menetelmää käytetään myös semanttisen webin sovelluksissa. URI-tunnisteiden syntaksi kuitenkin rajoittaa tunnisteissa sallittujen merkkien joukkoa.

Iconclass-järjestelmässä käytetään luokkien yksilöivinä tunnisteina notaatioita. Myös viittaukset luokkien välillä esitetään notaatioita käyttäen. Notaatiolla on siis Iconclass-järjestelmässä samankaltainen tehtävä kuin URI-tunnisteella WWW-arkkitehtuurissa. Tästä syystä luokittelujärjestelmän notaatiot ovat luonnollinen lähtökohta ontologian käsitteiden URI-tunnisteiden muodostamiseen. Seuraavissa luvuissa käsitellään sekä notaatioiden että URI-tunnisteiden syntaksi ja tämän jälkeen esitellään rajoitteet, joita notaatioiden esittämiseen URI-tunnisteina liittyy.

4.2.1 Notaatioiden syntaksi

Kuten luvussa 3.3 esitettiin, Iconclass-luokittelujärjestelmän notaatiot koostuvat numeroista '0-9' ja kirjaimista 'A-I, K-Z'. Luvussa 3.4.4 käsiteltyä avain-laajennusta käyttävissä notaatioissa esiintyy lisäksi plusmerkki '+' ja sulkeet '()'.

Mikäli notaatiossa käytetään luvussa 3.4.1 esitettyä nimi sulkeissa -laajennusta, ei notaatiossa sallittujen merkkien joukko ole yksikäsitteisesti määritettävissä, sillä Iconclass-järjestelmä ei aseta rajoitteita sulkeisiin merkittävälle tekstile. Laajennus perustuu englanninkielisen kontrolloidun sanaston käyttöön, joten pääosa laajennusta käyttävistä notaatioista on sisältää vain latinalaisen aakkoston merkkejä. Notaatioissa voi kuitenkin esiintyä myös esimerkiksi diakriittisiä merkkejä (82A (BARON MÜNCHAUSEN)) ja erilaisia välimerkkejä (73D271 (Jo. 14:2)).

4.2.2 URI-tunnisteiden syntaksi

Ontologian URI-tunniste koostuu nimiavaruudesta ja paikallisnimestä. Nimiavaruus toimii tunnisteena koko ontologialle, ja paikallisnimeä käytetään käsitteiden yksilöimiseen ontologian sisällä. Mikäli nimiavaruuden URI-tunniste päättyy

/-merkkiin (*slash namespace*), liitetään käsitteen paikallisnimi suoraan URI-tunnisteen perään. Paikallisnimi on tällöin osa URI-tunnisteen polkua. Tämä menetelmä mahdollistaa yksittäisten käsitteiden haun ontologiasta ja soveltuu siksi käytettäväksi suurissa ja usein muuttuvissa ontologioissa. Muussa tapauksessa liitetään nimiavaruuden URI-tunnisteen loppuun ensin #-merkki (*hash namespace*) ja tämän perään käsitteen paikallisnimi, jonka muoto vastaa tällöin resurssin sisäiseen viittaukseen käytettävää URI-tunnisteen loppuun liitettävää fragmenttiosaa. Tämä tapa soveltuu pienempiin ja harvoin muuttuviin ontologioihin, jossa tieto koko ontologian sisällöstä voidaan välittää kerralla. (Berners-Lee et al., 2004; Klyne & Carrol, 2004; Miles et al., 2006)

Koska Iconclass on keskitetysti hallintoitu luokittelujärjestelmä, tapahtuu muutoksia luokittelujärjestelmän rakenteeseen vain harvoin ja ennalta määritellyn prosessin kautta. Luokittelujärjestelmästä rakennettavaa ontologiaa voidaan pitää hyvin rajattuna ja suhteellisen staattisena kokonaisuutena. Toisaalta kymmeniä tuhansia luokkia sisältävä järjestelmä on varsin laaja. Tuotantokäytössä mahdollisuus yksittäisten käsitteiden hakuun ontologiapalvelimelta helpottaa ontologian käyttöä. Kehitystyössä ontologiaa käsitellään usein yhtenä kokonaisuutena, jolloin ontologia on tallennettu paikallisesti tiedostona. Tässä työssä paikallisnimelle käytetään URI-tunnisteen fragmenttiosaa vastaavaa muotoa. Ontologia voidaan kuitenkin kuitenkin helposti muuttaa paikallisnimen osana URI-tunnisteen polkua esittävään muotoon, sillä molempien muotojen rajoitteet on huomioitu paikallisnimeä muodostettaessa.

Ainoastaan seuraavia US-ASCII-merkistön merkkejä on sallittua käyttää sekä URI-tunnisteen polussa että fragmenttiosassa. (Berners-Lee et al., 2005):

- kirjaimia A-Z ja a-z
- numeroita 0-9
- merkkejä ! \$ & ' () * + , - . : ; = @ _

Lisäksi muiden merkkien esittämiseen voidaan käyttää prosentti-enkoodausta, jos-

sa esitetään merkki sen US-ASCII-arvoa vastaava heksadesimaalilukuna. Esimerkiksi merkintä %20 merkitsee välilyöntiä.

4.2.3 Rajoitteet notaation esittämiselle URI-tunnisteena

Iconclass-luokittelujärjestelmän runkoon kuuluvat notaatiot ovat sellaisenaan esitettävissä URI-tunnisteina, sillä ne sisältävät ainoastaan US-ASCII-merkistöön kuuluvia kirjaimia ja numeroita. Nimi sulkeissa -laajennusta käyttävien notaatioiden esittäminen URI-tunnisteena on ongelmallista, koska laajennusta käyttävä notaatio saattaa sisältää merkkejä, joita ei voida esittää URI-tunnisteessa ilman prosentti-enkoodauksen käyttöä.

Prosentti-enkoodauksen käyttöön liittyy kuitenkin ongelmia. Koodattu URI-tunniste saattaa viitata samaan RDF-resurssiin kuin toinen tunniste, jota ei ole koodattu. Koodaus tulisi aina purkaa viittausta selvitetessä, mutta URI-tunnistetta tulkitsevat ohjelmistot eivät käytännössä välttämättä menettele näin. Tästä syystä prosentti-enkoodauksen käyttöä RDF-resurssien URI-tunnisteissa suositellaan välttämään. (Klyne & Carrol, 2004) Esimerkiksi YSO:n käsitteen kissa URI-tunniste voidaan esittää muun muassa muodoissa `http://yso.fi/YSO#kissa` ja `http://yso.fi/YSO#%6B%69%73%73%61`. URI-tunnistetta tulkitsevan ohjelmiston on purettava prosentti-enkoodaus selvittääkseen, että kysymyksessä on sama URI-tunniste.

URI-tunniste on tarkoitettu ensisijaisesti koneellisesti käsiteltäväksi. Useat semanttisen webin työkalut vaativat kuitenkin vielä ihmiseltä URI-tunnisteiden tulkitsemista ja tunnistamista. Ihmiskäyttäjän kannalta olisikin edullista, mikäli URI-tunniste olisi mahdollisimman helposti luettava ja muistettava. Iconclass-järjestelmän notaatiot määräytyvät luokittelujärjestelmän hierarkkisen rakenteen mukaisesti, joten niistä muodostetut URI-tunnisteet eivät ole kovin mnemonisia ja helposti muistettavia. Luokittelujärjestelmää ammatikseen hyödyntävä käyttäjä oppii kuitenkin muistamaan yleisimmin tarvitsemansa notaatiot.

Ihmiskäyttäjän kannalta prosentti-enkoodaus heikentää URI-tunnisteen luettavuutta ja muistettavuutta. Koodauksessa käytettävät heksadesimaaliluvut muistuttavat

syntaksiltaan läheisesti Iconclass-järjestelmän notaatioissa käytettävää muotoa. Tästä syystä prosentti-enkoodauksen käyttö huonontaisi merkittävästi notaation luetavuutta URI-tunnisteesta.

4.2.4 URI-tunnisteiden sarjallistaminen XML-muotoon

Semanttisen webin sovellukset käyttävät usein RDF-kuvausten sarjallistamista XML-muotoon tietojen tallentamiseen ja välittämiseen. Sarjallistettujen XML-elementtien nimet muodostetaan URI-tunnisteiden paikallisnimistä. Tästä syystä URI-tunnisteiden luomisessa on otettava huomioon myös XML-syntaksin elementtien nimeämiselle asettamat rajoitteet.

XML-elementtien nimien on noudatettava Qualified Name -muodon paikallisnimelle asettamia rajoituksia (Bray et al., 1999):

- Nimen ensimmäisen merkin on oltava kirjain tai alaviiva ('_')
- Nimessä voi esiintyä kirjaimia, numeroita, pisteitä ('.'), yhdysmerkkejä ('-'), alaviivoja ('_') ja diakriittisiä merkkejä (*combining characters and extenders*)
- Nimen alussa ei saa esiintyä kirjaimia xml (pien- tai suuraakkosin kirjoitettuna)

WWW:n arkkitehtuurin kuvaava W3C:n suositus vaatii, että sovelluksen, joka käyttää Qualified Name -muodossa esitettäviä tunnisteita resursseille, on määritettävä menetelmä Qualified Name -tunnisteiden muuntamiseksi URI-tunnisteiksi. Suosituksessa ei kuitenkaan kuvata menetelmää, jolla muutos tulisi suorittaa. (Berners-Lee et al., 2004)

4.2.5 Rajoitteet notaation esittämiselle XML-elementin nimenä

Iconclass-luokittelujärjestelmän notaatioiden ensimmäisenä merkkinä on aina numero, joka ei ole sallittu Qualified Name -muodossa elementin paikallisnimen ensimmäisenä merkkinä. Ensimmäisen merkin jälkeen numeroiden käyttö paikallisnimessä on sallittua. Paikallisnimen muodostaminen notaatiosta edellyttää siis kirjaimella tai alaviivalla alkavan etuliitteen liittämistä notaation eteen.

Avain- tai nimi sulkeissa -laajennuksia käyttävät notaatiot sisältävät useita sellaisia merkkejä, joita ei voida esittää XML-elementin paikallisnimessä. Paikallisnimien muodostamiseksi näistä notaatioista on määritettävä korvauskäytäntö, jolla notaatiot muutetaan soveltuvaan muotoon.

4.2.6 ICONCLASS-ontologian URI-tunnisteet

ICONCLASS-ontologian käsitteiden URI-tunnisteet perustuvat luokittelujärjestelmän notaatioihin. Ontologian nimiavaruutena on <http://seco.tkk.fi/onto/iconclass/iconclass#>. Käsitteiden paikallisnimet muodostetaan seuraavasti. Paikallisnimen alkuun liitetään etuliite `not_`, mikäli kyse on luokittelujärjestelmän päähierarkiaan kuuluvasta notaatiosta. Jos notaatio esittää avain-laajennukseen kuuluvaa käsitettä, käytetään etuliitettä `key_` erottamaan avaimet luokittelujärjestelmän muista notaatioista. Etuliitteen käytöllä varmistetaan myös, että URI-tunniste voidaan sarjallistaa XML-muotoon. Etuliitteen jälkeen seuraa varsinainen aakkosnumeerinen notaatio. Mikäli notaatio sisältää välilyöntejä, ne poistetaan URI-tunnistetta muodostaessa. Iconclass-järjestelmä sallii notaation ryhmittelyn välilyönneillä luettavuuden parantamiseksi, mutta sillä ole vaikutusta notaation merkitykseen. Esimerkiksi notaatiolle 34B12 "kissa" tulee näin URI-tunniste http://seco.tkk.fi/onto/iconclass/iconclass#not_34B12.

Mikäli notaatiossa käytetään nimi sulkeissa -laajennusta, liitetään sulkeisiin kirjoitettua nimeä vastaava osa notaation perään kahdella alaviivalla erotettuna. Sulkeisiin liitetyn osan käsittelyssä noudatettava erillistä korvauskäytäntöä paikallisnimeen soveltumattomien merkkien poistamiseen. Korvauskäytäntö perustuu diakriittisten merkkien korvaamiseen lähimmillä US-ASCII-merkistöstä löytyvillä vastineillaan sekä muiden paikallisnimeen soveltumattomien merkkien korvaamiseen alaviivalla. Ainoa Kalevalan tarustoa edustava notaatio Iconclass-järjestelmässä 82A(LEMMINKÄINEN) saa URI-tunnisteen http://seco.tkk.fi/onto/iconclass/iconclass#not_82A__LEMMINKAINEN.

Avain-laajennukseen kuuluvilla käsitteillä etuliitettä `key_` seuraa sen käsitteen notaatio, johon kyseinen avain liittyy. Tunnisteen päättää avaimen numerokoo-

di, joka erotetaan notaatioista kahdella alaviivalla. Notaatioon 34 ”ihminen ja eläin” liitettylle avaimelle (+1) ”eläimen ruokkiminen” saadaan URI-tunniste http://seco.tkk.fi/onto/iconclass/iconclass#key_34__1.

4.3 Luokittelun merkityksen esittäminen

Koska Iconclass on kuvataiteeseen keskittynyt erikoisontologia, tarvitaan useissa sovelluskohteissa yhteyksiä ICONCLASS-ontologian ja yleisten ontologioiden kuten YSO:n välillä. Vaikka yleiset ontologiat eivät sisällä vastineita kaikille tietyn aihealueen erikoiskäsitteille, ne tarjoavat kuitenkin mahdollisuuden yhdistää eri aihealueiden ontologioita toisiinsa yhteisen yläkäsitteistön kautta ja näin kuvata käsitteiden merkitystä ja niiden välisiä suhteita myös eri aihealueiden välillä (Hyvönen et al., 2005).

4.3.1 Käsitevastaavuudet

Yksinkertaisimmillaan ICONCLASS-ontologian käsitteiden merkitystä voidaan kuvata liittämällä käsitteeseen suhde toisessa ontologiassa olevaan merkitykseltään samansisältöiseen käsitteeseen. YSO:n ja muiden ontologioiden välillä käsitteiden väliset yhteydet voidaan kuvata `yso:ontologyLink` suhteella, joka yhdistää YSO:ssa määritellyn käsitteen toisessa ontologiassa määriteltyyn merkitykseltään vastaavaan käsitteeseen. Esimerkiksi notaatioista 34B12 muodostetun ontologian käsitteen merkitystä voidaan ilmaista yhdistämällä se `yso:ontologyLink` suhteella YSO:n käsitteeseen `yso:kissa`, koska molemmat käsitteet merkitsevät kissaa.

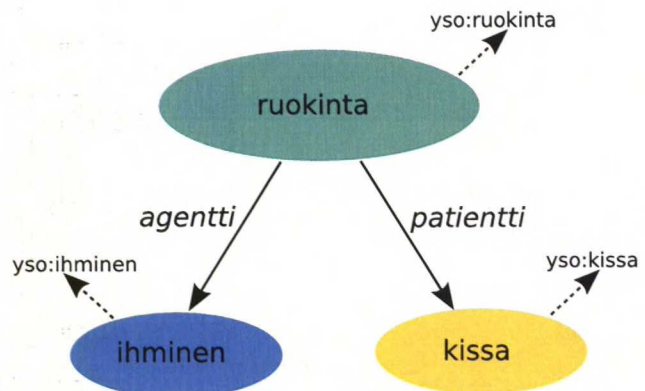
Kuvattaessa ICONCLASS-ontologian käsitteiden merkitystä YSO:n avulla haasteena on löytää YSO:sta mahdollisimman lähellä olevat vastineet käsitteille. Käsitevastaineiden haussa voidaan hyödyntää Iconclass-järjestelmän aakkosellista indeksiä, joka sisältää kullekin notaatiolle joukon siihen liittyviä asiasanoja. Mikäli asiasanaa vastaava käsite löytyy YSO:sta, voidaan ontologioiden välille muodostaa suhde. Menetelmän heikkoutena kuitenkin on, että saman sanan on löydettävä sekä Iconclass-järjestelmän indeksistä että YSO:sta. Esimerkiksi ruoan valmistusta grillaamalla merkitsevän notaation 41C2512 yhteys käsitteeseen `yso:grillit` jää tällä

menetelmällä löytymättä, koska indeksissä luetellaan ruokaan ("ravinto, ruoka"), suoritettavaan toimintaan ("keittäminen, ruoanvalmistaminen, paistaminen, paahdaminen") ja apuvälineisiin ("parila, halstari") liittyviä asiasanoja, mutta ei mainita itse grilliä.

Myös asiasanojen käännökset tuottavat ongelmia, koska käännökseen saattaa liittyä merkityksiä, jotka eivät kuulu alkuperäiseen asiayhteyteen. Esimerkiksi pappien käyttämiä liturgisia vaatteita kuvaavaan notaatioon 11Q71462 on liitetty englanninkielinen asiasana "robe" merkityksessä "kaapu, viitta". Aakkosellisen indeksin suomennoksessa asiasanaan on liitetty myös sen kolmas merkitys "kylpytakki". Näin kylpytakkia etsiessä saattaa tulokseksi tulla myös muunlaisia kaappuja ja viittoja.

4.3.2 Tapahtumakuvailu

Osa syvemmällä luokittelujärjestelmän hierarkiassa olevista notaatioista kuvaa yksittäisiä tapahtumia. Näissä tapauksissa pelkät asiasanoista muodostetut käsitevas-
tineet ontologioiden välillä eivät välitä tietoa tapahtuman sisäisestä semantiikasta. Käyttämällä samantapaista tapahtumaperustaista annotointiskeemaa kuin esimerkiksi KulttuuriSammossa voidaan tapahtumakehyksen avulla kuvailla tarkemmin tapahtumaan liittyvä toiminta, toimija, väline, kohde ja tavoite (Salminen, 2006, s. 29). Tapahtuma on Iconclass-luokittelujärjestelmässä kuvattu notaation sanallisessa määritteessä, joten tapahtumakuvaailujen muodostaminen näistä edellyttää käsityötä.



Kuva 4.3: Esimerkki tapahtumaperusteisen annotoinnin käytöstä Iconclass-järjestelmässä. Notaatio 34B12(+1) voidaan esittää ruokintatapahtuman ilmentymänä, jossa ruokinnan kohteena on yksilö luokasta yso:kissa ja tapahtuman agenttina on yso:ihminen luokkaan kuuluva yksilö

Tapahtumaa kuvaavia notaatioita voidaan muodostaa myös avain-laajennusta käyttäen. Tällä tavoin muodostetuissa notaatioissa tapahtuman semanttisen rakenteen koneellinen tulkitseminen on mahdollista. Tämä edellyttää avaimen merkityksen tuntemista ja siihen perustuvien sääntöjen muodostamista, sillä eri yhteyksissä avaimella voidaan kuvata erilaisia tapahtumaan liittyviä rooleja. Esimerkiksi notaation 34 ”ihminen ja eläin” yhteydessä on määritelty avain (+1), joka merkitsee eläimen ruokintaa. Kyseessä siis on ruokintatapahtuma, jonka kohteena on eläin ja toimijana on ihminen. Notaatio itsessään määrittelee toiminnan kohteen tarkemmin, esimerkiksi notaatiossa 34B12(+1) kyse on kissan ruokinnasta. Kuva 4.3 esittää esimerkin tämän notaation kuvaamisesta tapahtumaperusteisella annotoinnilla. Tapahtumaa kuvataan kehyksellä, joka liittää toisiinsa tapahtuman osallistujat ja näiden roolit tapahtumassa.

4.4 Yhteenveto

Iconclass-luokittelujärjestelmän hierarkkinen perusrakenne soveltuu esitettäväksi SKOS Core -mallia käyttäen. Luokittelujärjestelmän esittäminen laajennuksineen edellyttää kuitenkin SKOS Core -mallin laajentamista eri käsitetyyppien notaatioiden ja avaimien erottamiseksi toisistaan.

Ontologian käsitteiden URI-tunnisteet voidaan muodostaa notaatioista, joita Iconclass-järjestelmässä käytetään luokkien tunnisteina. Tunnistetta muodostettaessa on huomioitava URI-tunnisteen syntaksin ja XML-serialisoinnin asettamat rajoitteet, sillä URI-tunniste on syntaksiltaan Iconclass-järjestelmän notaatiota rajoitetumpi.

Ontologian käsitteiden merkitystä voidaan kuvata määrittämällä käsitteelle vastineita muissa ontologioissa kuten YSO:ssa. Osa luokittelujärjestelmän notaatioista kuvaa tapahtumia, joiden tarkka kuvailu edellyttäisi tapahtumakehysten käyttöä. Tapahtumakehysten avulla voidaan ilmaista, millaisessa roolissa käsite liittyy tapahtumaan.

5 ICONCLASS-ONTOLOGIA

Tässä luvussa esitellään ICONCLASS-ontologian esittämiseen käytetty SKOS Core -mallin laajennus. Tämän jälkeen käsitellään muunnoksen käytännön toteutusta. Lopuksi tarkastellaan ICONCLASS-ontologian ja YSO:n yhdistämiseen käytettyjä menetelmiä.

5.1 ICONCLASS-ontologian malli

5.1.1 SKOS Core -mallin puutteet

SKOS Core -malli pyrkii nimensä mukaisesti tarjoamaan mahdollisimman yksinkertaisen mallin tiedon jäsentämiseen. Tästä johtuen rakenteeltaan monimutkaisia malleja, kuten Iconclass-luokittelujärjestelmää, ei voida kuvata täysin SKOS Core -mallia käyttäen. Merkittävimpiä puutteita olivat tarve kahden toisistaan erillisen käsitetyypin, notaatioiden ja avainten, esittämiseen sekä näiden välisten suhteiden erottamiseen samaa tyyppiä olevien käsitteiden välisistä suhteista.

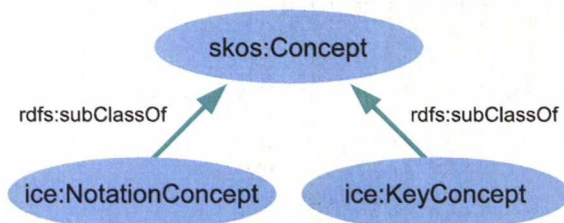
SKOS Core -mallissa on kuitenkin varauduttu näihin puutteisiin määrittelemällä malli helposti laajennettavaksi. Malliin voidaan määritellä uusia ominaisuuksia ja luokkia olemassa olevista määrittelyistä perimällä. Näin mallia voidaan tarvittavilta osilta täsmentää säilyttäen silti yhteensopivuus yleisen SKOS Core -mallin kanssa (Miles et al., 2005).

5.1.2 ICONCLASS-laajennus

Iconclass-luokittelujärjestelmän esittämistä varten SKOS Core -mallia laajennettiin kahdella uudella luokalla ja yhdellä ominaisuudella. Tässä luvussa esitellään laajennuksen semanttinen rakenne. Liite 1 esittää tätä vastaavan laajennuksen teknisen määrittelyn. skos:Concept-luokalle määritettiin uudet aliluokat ice:NotationConcept ja ice:KeyConcept notaatioiden ja avainten esittämistä varten kuvan 5.1 mukaisesti. Nimiavaruuden tunniste ice merkitsee SKOS Core -mallin ICONCLASS-laajennusta erotuksena varsinaisesta ICONCLASS-ontologiasta, jonka nimiavaruutta merkitään tunnisteella ic.

Uudet luokat mahdollistavat notaatioiden ja avainten erottamisen toisistaan. Samalla säilyy mahdollisuus käyttää `skos:Concept`-luokalle määritettyjä `skos:broader`, `skos:narrower` ja `skos:related` ominaisuuksia keskenään samantyyppisten käsitteiden välisen semanttisen rakenteen kuvaamiseen.

```
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix ice: <http://seco.tkk.fi/onto/iconclass/iconclass-skos-ext#>
```



Kuva 5.1: SKOS Core -mallin laajennus kahdella uudella luokalla notaatioille ja avaimille

Lisäksi määritettiin `ice:hasKeyConcept`-ominaisuus, jolla liitetään luokittelujärjestelmän runkoon kuuluva notaatio siihen avainpuuhun, johon kuuluvia avaimia voidaan kyseisen notaation yhteydessä käyttää. Tämä ominaisuus on esitetty kuvassa 5.2. Yhdessä laajennuksessa määritettyjen uusien luokkien kanssa tämä ominaisuus mahdollistaa avain-laajennuksen ja luokittelujärjestelmän välisen suhteen tarkemman kuvaamisen, sillä ominaisuus on rajoitettu käytettäväksi `ice:NotationConcept` ja `ice:KeyConcept` luokkien välillä `rdfs:range` ja `rdfs:domain` rajoitteiden avulla.



Kuva 5.2: Luokittelujärjestelmän rungon ja avain-laajennuksen välinen suhde

5.2 Muunnos ontologiaksi

5.2.1 Vastaavuuksien määrittely

Iconclass-luokittelujärjestelmästä oli käytettävissä XML-muotoinen tiedosto, joka esitti luokittelujärjestelmän hierarkkisen rakenteen ja sisälsi notaatiot määritteineen sekä notaatioihin liittyvät aakkosellisen indeksin avainsanat. Luokittelujärjestelmän ja avain-laajennuksen rakenteille määritettiin taulukossa 5.1 esitetyt vastaavuudet XML-tiedoston tietoalkioiden ja ontologiamallin luokkien ja ominaisuuksien välillä.

Taulukko 5.1: Iconclass-luokittelujärjestelmän muunnos XML-muodosta ontologiaksi

Tietoalkio	Merkitys	Luokka / Ominaisuus
<subject notation="X">	Notaatio	ice:NotationConcept, jolla rdf:about=not_X ja skos:prefLabel=X.
<subject notation="X"> <subject key="Y">	Avain	ice:KeyConcept, jolla rdf:about=key_XY ja skos:prefLabel=XY.
<text language="Y">X	Määrite	skos:definition=X, jolla xml:lang=Y.
<subject notation="X"> <subject notation="Y">	Notaatioiden LT-suhde	ice:NotationConcept Y, jolla skos:broader=X.
<subject key="X"> <subject key="Y">	Avainten LT-suhde	ice:KeyConcept Y, jolla skos:broader=X.
<subject notation="X"> <subject notation="Y">	Notaatioiden ST-suhde	ice:NotationConcept X, jolla skos:narrower=Y.
<subject key="X"> <subject key="Y">	Avainten ST-suhde	ice:KeyConcept X, jolla skos:narrower=Y.
<subject notation="X"> <xref-list> <xref>Y	Notaatioiden RT-suhde	ice:NotationConcept X, jolla skos:related=Y.
<subject key="X"> <xref-list> <xref>Y	Avainten RT-suhde	ice:KeyConcept X, jolla skos:related=Y.
<subject notation="X"> <subject key="Y">	Avaimen määrittely	ice:NotationConcept X, jolla ice:hasKeyConcept=Y.

5.2.2 Luokittelujärjestelmän laajennusten käsittely

Luokittelujärjestelmän laajennukset huomioitiin muunnoksessa seuraavasti:

Nimi sulkeissa -laajennus ja siihen liittyvät kontrolloidut sanastot voidaan mallintaa samalla tavalla kuin tavalliset notaatiot, koska sanastossa määritellyt notaatiot liittyvät LT- ja ST-suhteilla suoraan luokittelujärjestelmän runkoon, kuten luvussa 3.4.1 todettiin. Näin ollen laajennuksen rakenne ei tarvitse erikoiskäsittelyä muunnoksessa. URI-tunnisteet laajennuksen käsitteille muodostetaan luvussa 4.2.6 esitetyllä menettelyllä, joka muuttaa notaation URI-tunnisteen paikallisnimeksi soveltuvaan muotoon.

Laajennuksen yhteydessä on mahdollista käyttää myös vapaata asiasanoitusta, joka ei ole varsinaisesti osa luokittelujärjestelmää ja jää myös ontologian ulkopuolella. Annotoinnissa vapaata asiasana voidaan liittää annotaatioon literaaliarvona.

Kaksoiskirjaimet-laajennus määrittää luvussa 3.4.2 esitetyllä tavalla kaksi rinnakaista alipuuta, joiden merkitys on keskenään vastakohtainen tai muuten käänteinen. Laajennus noudattaa luokittelujärjestelmän hierarkkista rakennetta, joten se voidaan mallintaa tavallisten notaatioiden tapaan.

Koska laajennuksen määrittämien alipuiden välisen suhteen tyyppiä ei ole eksplisiittisesti ilmaistu luokittelujärjestelmässä, suhdetta ei ole esitetty myöskään ontologiassa. Esimerkiksi vastakohtaisuutta merkitsevän suhteen määrittäminen kaikkien kaksoiskirjaimet-laajennuksen alipuiden välille olisi virheellinen menettelytapa, koska laajennus ei välttämättä muuta käsitteen merkitystä täysin vastakohdakseen, vaan saattaa kääntää sen merkityksen vain osittain.

Rakenteellinen numerointi -laajennus määrittää menetelmän, jolla notaatiohierarkiaa voidaan laajentaa luvussa 3.4.3 kuvattujen tiettyjä henkilöhahmojen ryhmiä esittävien luokkien yhteydessä. Laajennus on luonteeltaan lähinnä luokittelujärjestelmän tiettyjen osien hallintaan liittyvä ohjeistus (Anon, 2005a), koska laajennuksen käyttöä ei voida havaita muuten kuin lukemalla luokittelujärjestelmästä, että kyseisen notaation muodostamisessa on käytetty laajennusta.

Koska rakenteellinen numerointi -laajennus on käytettävissä vain tiettyjen tarkasti määriteltyjen luokkien yhteydessä, eikä tieto laajennuksen käytöstä tarjoa käyttäjälle mitään notaation merkityksen tulkintaan vaikuttavaa informaatiota, tuotetaan muunnoksen yhteydessä laajennuksen avulla muodostettavia notaatioita vastaavat `ice:NotationConcept` käsitteet ontologiaan valmiiksi.

Avaimet esitetään edellä luvussa 5.1.2 esitetyn SKOS Core -mallin ICONCLASS-laajennuksen mukaisesti omina käsitehierarkioinaan. Sisäiseltä rakenteeltaan avain-laajennus on luokittelujärjestelmän rungon kaltainen sisältäen hierarkisia `skos:broader` ja `skos:narrower` sekä assosiatiivisia `skos:related` suhteita. Luokittelujärjestelmän rungossa mahdollisuus avain-laajennuksen käyttöön on ilmaistu liittämällä notaatio `ice:hasKeyConcept` suhteella siihen avainhierarkiaan, jota kyseisen notaation alipuun yhteydessä on mahdollista käyttää.

Avain-laajennuksen osana olevat rekursiivisten avainten avulla muodostetut alipuut, joita Iconclass-järjestelmässä kutsutaan avaimen jonoksi, laajennetaan muunnoksen yhteydessä avainhierarkian alipuihin `ice:KeyConcept` käsitteiksi. Tämä menettely valittiin koska rekursiivista avainta voidaan käyttää vain eräiden ennalta määrättyjen avainten yhteydessä. Rakenteellinen numerointi -laajennuksen tapaan tämä laajennus ei vaikuta avaimen merkityksen tulkitsemiseen.

Annotoinnissa avain-laajennuksen sisältäviä käsitteitä voidaan muodostaa dynaamisesti avainlaajennuksen sovellusalueen sallimissa rajoissa. Tämä tapahtuu liittämällä annotaatio `ice:hasKeyConcept` suhteella siihen avain-laajennuksen käsitteeseen, jota annotaation esittämässä notaatiossa on käytetty.

5.2.3 Muunnoksen toteutus

Koska SKOS Core -malli voidaan esittää RDF/XML-muodossa, toteutettiin ontologian muodostaminen XML-pohjaisten tiedostomuotojen välisiin muunnoksiin hyvin soveltuvalla XSLT-muunnoksella. Luokittelujärjestelmän laajennusten käsitteilyn vuoksi muunnoksen suorittamiseen käytettävä säännöstö muodostui varsin laajaksi. Kaikkiaan muunnossäännöt koostuvat viidestä XSL-tiedostoista joista yksi

vastaa luokittelujärjestelmän runkorakenteen läpikäynnistä, toinen URI-tunnisteiden muodostamisesta notaatioista, kolmas SKOS Core -mallin mukaisten luokkien ja ominaisuuksien tuottamisesta, neljäs ja viides tiedosto käsittelevät luokittelujärjestelmän laajennuksia.

Luokittelujärjestelmän muunnoksen lisäksi Iconclass-järjestelmän aakkosellinen indeksi muunnettiin SKOS Core -mallin mukaiseen muotoon. Koska indeksillä ei ole sisäistä hierarkkista rakennetta, oli sen muunnos yksinkertaista. Indeksistä muodostettu ontologia sisältää asiasanoista muodostettuja `skos:Concept` käsitteitä ja `skos:related` suhteilla ilmaistuja viitteitä varsinaisen luokittelujärjestelmän käsitteisiin.

5.3 Iconclass-ontologian yhdistäminen muihin ontologioihin

ICONCLASS-ontologia yhdistettiin YSO:aan määrittelemällä ontologioiden välille käsitevastineita. Käsitevastineiden hakuun kokeiltiin kahta automaattista menetelmää.

Ensimmäinen menetelmä perustuu luvussa 4.3.1 esitettyyn Iconclass-järjestelmän aakkosellisen indeksin käyttöön. Aakkosellisen indeksin asiasanojen perusmuotoa verrattiin YSO:n käsitteiden nimikkeeseen perusmuotoon Sampo Yrjänäisen SeCo-ryhmässä kehittämän Logical Ontology Mapper -työkalun avulla. Työkalu hakee ontologioista käsitteet, joiden nimikkeillä on sama perusmuoto, ja luo löytyneille vastineille `owl:equivalentClass` suhteen YSO:n ja ICONCLASS-ontologian käsitteiden välille.

Toisena menetelmänä kokeiltiin Iconclass-notaatioiden suomenkielisten määritteiden analysointia Natural Language Processing (NLP) -tekniikkaa käyttävällä Poka-työkalulla (Vehviläinen et al., 2006), joka etsii ontologian käsitteitä tekstistä. Koska käytössä ollut Iconclass-järjestelmän suomennos sisälsi käännökset vain osalle määritteistä, voitiin tällä menetelmällä hakea käsitevastineita vain osalle ontologian käsitteistä. Tästä syystä yhdistämiseen käytettiin ensimmäistä menetelmää, koska aakkosellinen indeksi sisälsi suomenkielisen asiasanoituksen myös suurimmalle osalle niistä notaatioista, joiden määritettä ei ollut suomennettu.

Iconclass-järjestelmässä monissa notaatioissa tärkeässä osassa ovat henkilöt ja paikat. Siksi myös aakkosellisessa indeksissä on runsaasti viittauksia henkilöihin ja paikannimiin. YSO sisältää ainoastaan yleiskäsitteitä, joten myös ICONCLASS-ontologian ja YSO:n väliset käsitevastaavuudet kohdistuvat näihin molemmissa ontologioissa esiintyviin yleiskäsitteisiin. Vastaavuuksien määrittäminen henkilöille ja paikannimille edellyttäisi näiden esittämiseen tarkoitettun erikoisontologian käyttöä ja ICONCLASS-ontologian yhdistämistä sen kanssa.

6 SOVELLUSKOHTEITA

Tässä luvussa käsitellään niitä ICONCLASS-ontologian sovelluskohteita, joihin ontologiaa on käytetty tämän työn puitteissa FinnONTO-projektissa. Ensiksi esitellään menetelmä, jolla Valtion taidemuseon Iconclass-järjestelmää käyttäen luokiteltu aineisto liitettiin osaksi KulttuuriSampo-portaalia. Tämän jälkeen tarkastellaan ICONCLASS-ontologian tarjoamia mahdollisuuksia konesemanttista suositelua hyödyntäville sovelluksille. Esimerkkinä käytetään luokittelijan aputyökalu ICONCLASS-ONKI:a, jota kehitetään ONKI-ontologiapalvelimen selailu- ja hakukäyttöliittymän pohjalta.

6.1 KulttuuriSampo -portaali

6.1.1 Iconclass-aineisto

Valtion taidemuseosta KulttuuriSampoa varten käyttöön saatu aineisto sisälsi 456 teoksen Iconclass-luokittelun sekä teos- ja tekijätiedot Muusa-järjestelmästä (Anon, 2006b) CIDOC CRM -muodossa (Doerr, 2003) ja digitaaliset valokuvat teoksista. Aineistojoukon kokoa rajoitti ensisijaisesti Iconclass-luokitellun aineiston saatavuus, sillä suurinta osaa Valtion taidemuseon kokoelmista ei ole vielä kuvailtu Iconclass-järjestelmällä. Museoiden luokittelukäytäntöjen välisten eroavaisuuksien havaitsemiseksi menetelmää kokeiltiin myös Foto Marburg -instituutista saadulla vertailuaineistoilla, joka sisälsi Iconclass-luokittelun 224 teokselle.

6.1.2 Muunnosmenetelmä

KulttuuriSampo-portaali yhdistää useita eri lähteistä saatuja aineistokokonaisuuksia. Koska aineistojen sisällöt on kuvailtu toisistaan poikkeavilla menetelmillä, on SeCo-ryhmässä kehitetty Parser-ohjelma, jolla eri tavoin kuvaillut aineistot voidaan muuntaa KulttuuriSammon annotaatio skeeman mukaiseen muotoon. Valtion taidemuseon aineiston käsittelyä varten ohjelmaan laajennus, joka osaa tulkita Iconclass-luokittelua sekä Muusa-järjestelmästä saatuja teoksen metatietoja.

Kaikille luokittelussa käytetyille notaatioille ei löydy suoraa vastinetta ontologiasta. Teoksen liitetyissä notaatioissa saattaa olla käytetty esimerkiksi vapaata asiasa-

naa osana nimi sulkeissa -laajennusta tai notaatio voi olla lisätty Iconclass-järjestelmään äskettäin, jolloin sitä ei löydy vielä ontologian muodostamisessa käytetystä versiosta. Tästä syystä Iconclass-luokittelua tulkitessaan ohjelma pyrkii löytämään ICONCLASS-ontologiasta mahdollisimman tarkan vastineen teoksen kuvailumiseen käytetylle notaatiolle.


Ohjelma tulkitsee teoksen luokittelussa käytettyjä notaatioita seuraavaa iteratiivista algoritmia käyttäen. Algoritmi ottaa syötteenään teoksen luokittelussa käytetyn notaation. Algoritmi palauttaa notaatiota ICONCLASS-ontologiassa mahdollisimman tarkasti vastaavan käsitteen URI-tunnisteen tai virheilmoituksen, mikäli notaatioille ei löydy vastinetta ontologiasta.

1. Tarkista, sisältääkö syötteenä annettu notaatio avain-laajennusta. Mikäli laajennusta on käytetty, se erotetaan notaatiosta ja käsitellään erikseen.
2. Tarkista, löytyykö ontologiasta käsite, jonka nimike vastaa notaatiota. Mikäli notaatiota vastaava käsite löytyy, lopeta käsittely ja palauta ontologian käsitteen URI-tunniste.
3. Siirry notaatiosierarkiassa yksi taso ylöspäin: Jos notaatio päättyy nimi sulkeissa -laajennukseen, ylemmälle tasolle siirrytään poistamalla se notaatiosta. Muussa tapauksessa hierarkiassa yhtä tasoa ylempänä oleva notaatio saadaan poistamalla nykyisestä notaatiosta sen viimeinen merkki. Mikäli notaatio ei tämän jälkeen ole tyhjä merkkijono, siirry kohtaan 2.
4. Jos notaatio on tyhjä merkkijono, sille ei löytynyt vastinetta ontologiasta. Palauta virheilmoitus ja lopeta käsittely.

Avain-laajennusta tulkittaessa haetaan ontologiasta käsitettä, jonka yhteydessä kyseinen avain on määritelty. Ohjelma käyttää seuraavaa algoritmia, joka ottaa syötteenään luokittelujärjestelmän runkoon kuuluvan notaation ja siihen liitetyn avaimen. Algoritmi palauttaa avainta vastaavan ICONCLASS-ontologian käsitteen URI-tunnisteen tai virheilmoituksen, mikäli avainta ei ole määritelty ontologiassa syötteenä annetun notaation yhteydessä.

1. Tarkista, onko syötteenä annettua avainta määritelty ontologiassa syötteenä annettua notaatiota vastaavan käsitteen yhteydessä. Jos avain on määritelty, ontologiassa on `ice:hasKeyConcept` suhde notaatiota vastaavasta käsitteestä siihen avainhierarkiaan, jossa syötteenä annettua avainta vastaava käsite sijaitsee. Mikäli avain on määritelty, lopeta käsittely ja palauta avaimen URI-tunniste.
2. Siirry notaatioshierarkiassa yksi taso ylöspäin: Poista notaatiosta sen viimeinen merkki. Mikäli notaatio ei ole tyhjä merkkijono, liitä siihen alkuperäinen avain ja siirry kohtaan 2.
3. Jos notaatio on tyhjä merkkijono, avainta ei ole määritelty ontologiassa. Palauta virheilmoitus ja lopeta käsittely.

Esimerkiksi Akseli Gallen-Kallelan maalaus Kullervon kirous on luokiteltu käyttäen kuvassa 6.1 esitettyjä Iconclass-notaatioita. Teoksen luokittelusta löytyy esimerkkejä Iconclass-järjestelmän eri laajennusten käytöstä. Liite 2 esittää maalauksen Iconclass-luokittelusta muodostetun annotaation. KulttuuriSammon annotaatio skeemassa sisällönkuvailu esitetään teosta kuvaavana tilanteena, johon ontologian käsitteet liittyvät (Hyvönen et al., 2006). Iconclass-luokittelu yhdistetään teoksen tilannekuvailuun siten, että teoksen sisältöä kuvaavista ICONCLASS-ontolo-



81 (KALEVALA)	kirjalliset kertomukset ja tarut (Kalevala)
82A (KULLERVO)	kirjallisten henkilöahmojen nimet (miespuoliset kirjalliset henkilöahmot) (Kullervo)
13B42	kiroaminen
56E2	raivo, viha
31D112	nuori mies
31A231	seisova figuuri
31A25113	käsi kohotettuna nyrkissä
34B11	koira
25H15	metsä
25H11245	siirtolohkare
25G3 (PIHLAJA)	puut (pihlaja)
25G1 (+21)	runko

Kuva 6.1: Akseli Gallen-Kallelan maalaus Kullervon kirous ja sen Iconclass-luokittelu (Ateneumin taidemuseo)

gian käsitteistä luodaan KulttuuriSammon annotaatio-ontologiaan ilmentymät, jotka liitetään teoksen tilannekuvaukseen mapIC suhteella. Seuraavassa esitetään eräiden teokseen liittyvien notaatioiden vastineiden haku ontologiasta.

Notaatio 81 (KALEVALA) on määritelty nimi sulkeissa -laajennuksen sanastossa, joten sitä vastaava käsite `ic:not_81__KALEVALA` löytyy suoraan ontologiasta. Sen sijaan notaatiossa 82A (KULLERVO) on Kullervo on esitetty vapaata asiasanaa käyttäen, koska Kullervoa ei löydy laajennuksen sanastosta. Lähin ontologiasta löytyvä vastine tälle notaatiolle on `ic:not_82A__...`, joka merkitsee miespuolisia kirjallisia henkilöitä.

Avain-laajennusta käyttävälle notaatiolle 25G1(+21) haetaan ICONCLASS-ontologiasta erikseen käsitevastineet notaatiolle ja avaimelle. Näistä käsitteistä teokselle muodostetaan kuvan 6.2 mukainen annotaatio, jossa notaatio liitetään avaimeen `ice:hasKeyConcept` suhdetta käyttäen.

```
ann:not_25G1-1161007574839
  a ic:not_25G1 ;
  ice:hasKeyConcept ic:key_25G__21
```

Kuva 6.2: Iconclass-notaatio 25G1(+21) KulttuuriSammon annotaatiossa

KulttuuriSammon annotaatioskeema ja suosittelujärjestelmä perustuvat YSO:n käsitteiden käyttöön. Teoksen luokittelussa käytettyjä Iconclass-notaatioita vastaavat YSO:n käsitteet voidaan hakea ontologioiden välisten käsitevastineiden avulla. Koska suosittelujärjestelmän tulee toimia portaalissa reaaliaikaisesti, pidentää useiden käsitteiden välisten suhteiden läpikäyntiä edellyttävä käsitevastineiden haku suoritusajaa tarpeettomasti. Suosittelun nopeuttamiseksi YSO:sta haettiin kaikkien teoksen luokittelussa käytettyjen Iconclass-notaatioiden käsitevastineet, ja näistä muodostettiin YSO:n käsitteisiin perustuva tilannekuvailu, jota voidaan suoraan hyödyntää suosittelujärjestelmässä. Kuva 6.3 esittää esimerkin maalaukselle Kullervon kirous Iconclass-luokittelun pohjalta muodostetusta YSO:n käsitteisiin perustuvasta tilannekuvailusta.



Kuva 6.3: Esimerkki maalaukselle Kullervon kirous Iconclass-luokittelun perusteella muodostetusta YSO:n käsitteisiin perustuvasta KulttuuriSammon annotaatiooskeeman mukaisesta tilannekuvaailusta. Violetti nuoli esittää maalauksen ja tilanteen välisen hasScene suhteen. Oranssit nuolet esittävät tilanteen ja siihen liittyvien käsitteiden väliset elementtiInstance suhteet. Mustat nuolet ovat käsitteiden tyyppin määrittäviä rdf:type suhteita. Katkoviivat liittävät käsitteisiin niiden merkitystä kuvaavan literaaliarvon

6.1.3 Tulosten arviointi

Käsitellyn aineiston toimivuutta KulttuuriSammon haku- ja suosittelujärjestelmien kannalta arvioitiin laskemalla aineiston Iconclass-luokittelulle YSO:sta löytyneiden käsitevastineiden määrä. Taulukko 6.1 esittää käsiteltyjen notaatioiden lukumäärän KulttuuriSampoa varten käsitellyssä Valtion taidemuseon aineistossa ja Foto Marburg -instituutin vertailuaineistossa sekä notaatiolle suoraan ja notaation yläluokalle löytyneiden vastineiden lukumäärän.

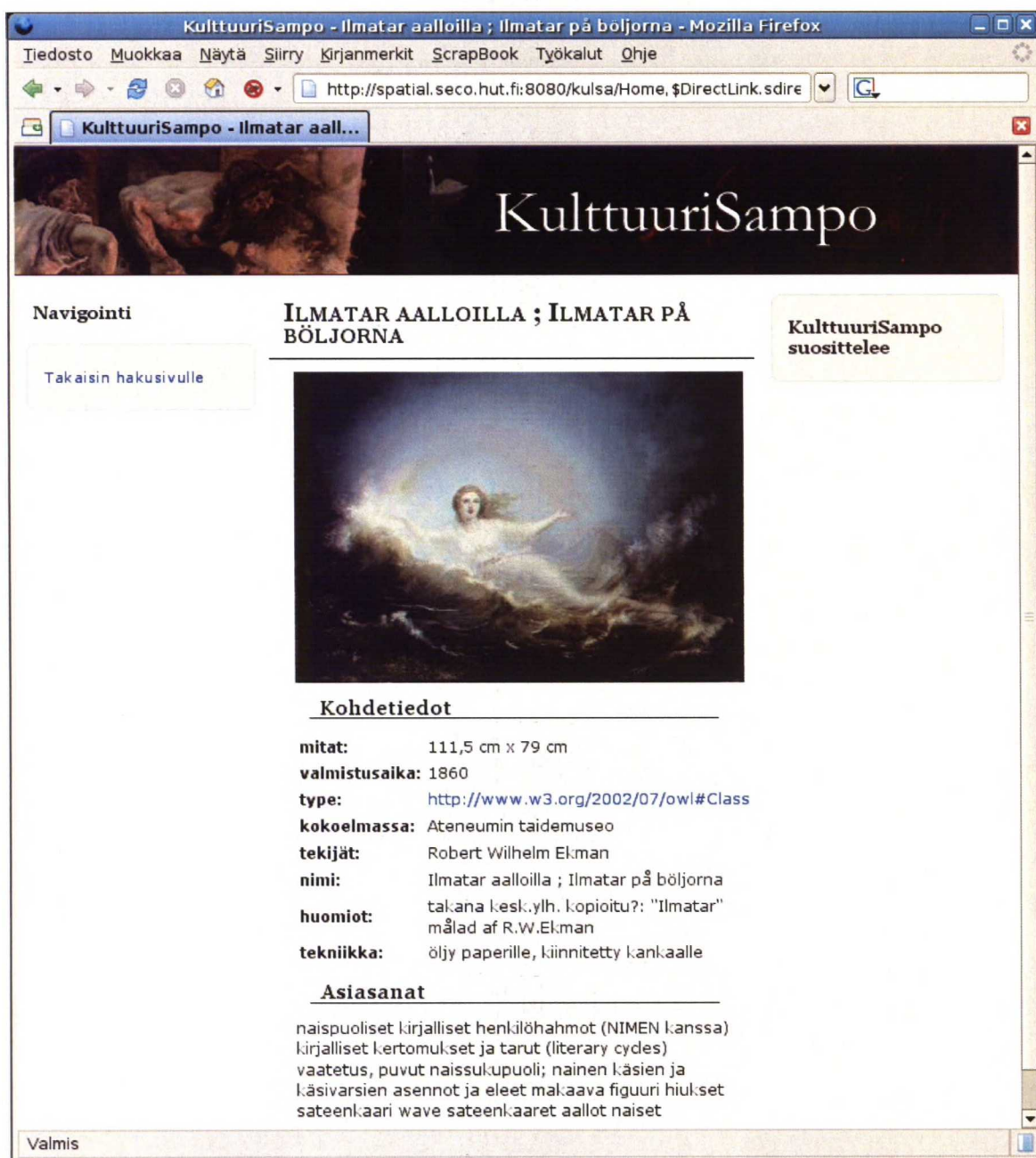
Taulukko 6.1: Iconclass-luokitellun aineiston käsitevastineet YSO:n kanssa

Aineisto	Teoksia	Notaatioita	Suora vastine	Yläluokan vastine	Ei vastinetta
VTM	456	2676	2206 (82,4 %)	333 (12,4 %)	137 (5,1 %)
Foto Marburg	224	532	356 (66,9 %)	114 (21,4 %)	62 (11,7 %)

Suurimmalle osalle notaatioista löytyi suora vastine molemmista aineistoista. Yläluokan vastineita haettiin niille notaatioille, joille ontologiasta ei löytynyt suoraa vastinetta. Suurimmassa osassa näistä oli käytetty nimi sulkeissa -laajennusta vapaan asiasanoituksen kanssa. Yläluokan vastineiden suurempi määrä Foto Marburg -aineistossa kertoo nimi sulkeissa -laajennuksen runsaasta käytöstä tässä aineistossa. Osalle notaatioista ei löytynyt vastinetta ontologiasta. Suurimpana syynä käsitevastineen puuttumiseen oli se, että notaatio viittasi sellaiseen luokittelujärjestelmän osaan, jota ei ole suomennettu.

Molemmista aineistoista löydettiin käsitevastine lähes yhdeksälle notaatiolle kymmenestä. Tämä tulos on varsin hyvä, sillä yksittäinen teos on luokiteltu yleensä useammalla kuin yhdellä notaatiolla. Käsitevastineiden puuttuminen yksittäisiltä notaatioilta merkitsee, että joidenkin teosten luokittelua ei voida kuvata täydellisesti, mutta valtaosa siitä voidaan esittää menetelmän avulla.

KulttuuriSampoon tuotiin edellä esiteltyä menetelmää käyttäen yhteensä 456 Iconclass-järjestelmällä luokiteltua teosta (pääasiassa maalauksia ja grafiikkaa) Val-

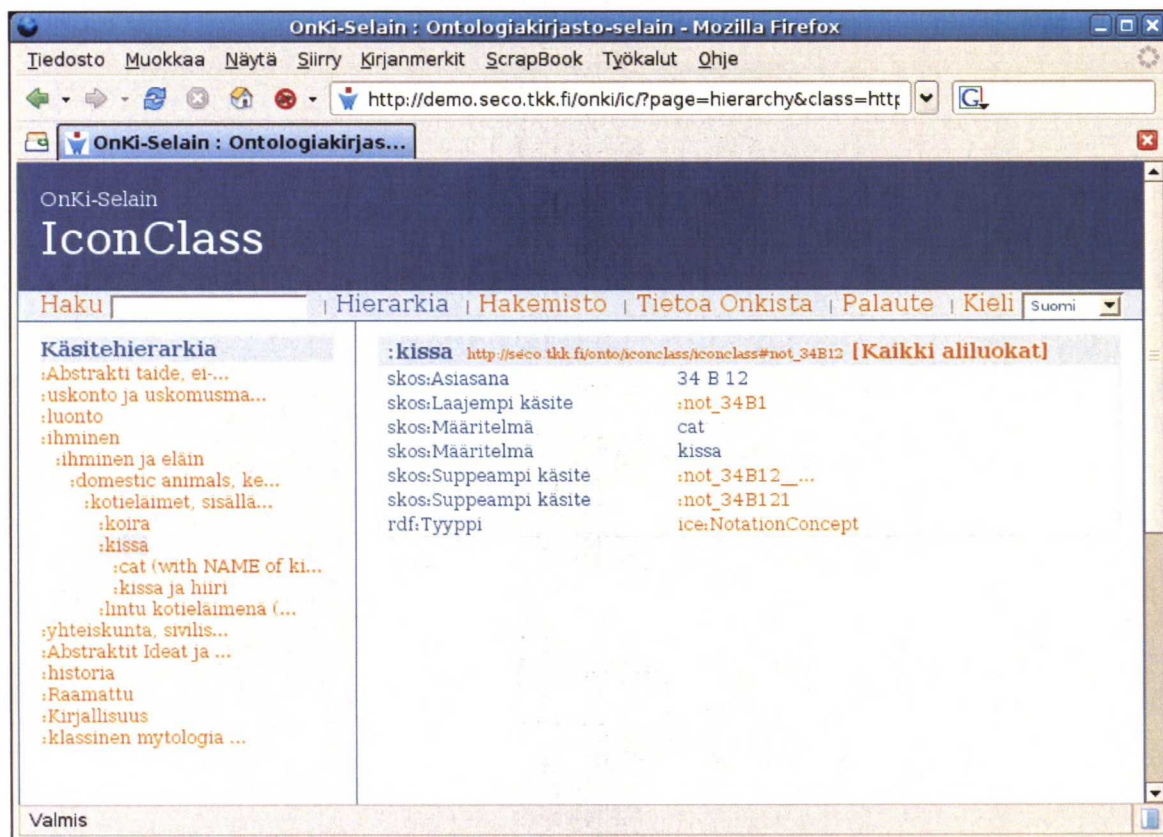


Kuva 6.4: Esimerkki Iconclass-luokitellun maalauksen kohdesivusta KulttuuriSampo-portaalissa

tion taidemuseon kokoelmista. Kuva 6.4 esittää maalauksen tietosivun Kulttuuri-Sampo-portaalissa.

6.2 Luokittelijan apuväline ICONCLASS-ONKI

Tässä luvussa selvitetään, miten ICONCLASS-ontologiaa ja konesemanttista päätelyä voitaisiin käyttää apuvälineenä taideteosten luokittelussa. Tavoitteena on helpottaa luokittelijan työtä tarjoamalla hänelle nykyistä parempia työvälineitä. Tämän työn yhteydessä ICONCLASS-ontologia liitettiin FinnONTO-projektissa aiemmin kehitettyyn ONKI-selaimeen (Komulainen et al., 2005). Kuva 6.5 esittää ICONCLASS-ontologian ONKI-selaimessa. Ontologiaselain ei kuitenkaan yksinään tarjoa riittävästä tukea luokittelijan työhön. Seuraavassa tarkastellaan, miten ontologiaselainta tulisi kehittää, jotta sitä voitaisiin käyttää luokittelijan apuvälineenä. Tarkastelu perustuu Lüscherin (2004, s. 24-26) Pro gradu -tutkielmassaan kuvaamiin luokittelutyön haasteisiin.



Kuva 6.5: ICONCLASS-ontologia ONKI-selaimessa

6.2.1 Oikeiden notaatioiden löytäminen

Luokittelu määrittelee tarjonnan hakujen tekijälle. Luokittelijan tulisi löytää teokselle sen sisältöä parhaiten vastaavat notaatiot, jotta teos löytyisi hauissa oikeissa asiayhteyksissä.

Oikeiden notaatioiden valintaa voidaan pyrkiä helpottamaan semanttisen suosittelun avulla (Viljanen, 2006). Suosittelun tavoitteena voi olla joko luokittelun tarkentaminen tarjoamalla teosta luokittelijan valitsemaa notaatiota mahdollisesti tarkemmin kuvaavia notaatioita tai teoksen kuvailun rikastaminen ehdottamalla valittujen notaatioiden kanssa usein yhdessä esiintyviä notaatioita. Suositukset voivat perustua joko yksittäiseen teokseen liitettyyn notaatioon tai useampaan notaatioon yhdessä. Taulukko 6.2 esittää yhteenvedon suosittelun tavoitteista ja perusteista.

Taulukko 6.2: Semanttisen suosittelun tavoitteet ja perusteet.

Suosittelun tavoite	Suosittelun peruste
luokittelun tarkentaminen	yksittäinen teoksen kuvailuun käytetty notaatio
kuvailun rikastaminen	joukko teoksen kuvailuun käytettyjä notaatioita

Yksittäiseen notaatioon perustuva suosittelu soveltuu tarkentavien notaatioiden ehdottamiseen. Tämä voidaan yksinkertaisimmillaan toteuttaa luokittelujärjestelmän sisäisiä ristiviitteitä hyödyntämällä. Ristiviitteet eivät yksinään kuitenkaan ole riittävän kattava verkko suosittelujärjestelmän pohjaksi. Esimerkiksi ihmistä merkitsevät notaatiot ovat luokittelujärjestelmässä hajallaan, eikä kaikkia niiden välisiä suhteita ole merkitty ristiviitein.

Luokittelujärjestelmän ristiviitteiden lisäksi myös aakkosellista indeksiä voidaan hyödyntää suosittelussa. Indeksi tarjoaa mahdollisuuden hakea muita notaatioita, joihin on liitetty samoja asiasanoja kuin luokittelijan valitsemaan notaatioon. Esimerkiksi notaatioon 3 ”ihminen” on aakkosellisessa indeksissä liitetty asiasanaksi ihminen. Sama asiasana on indeksissä liitetty myös taulukossa 6.3 esitettyihin no-

taatioihin, joten tässä tapauksessa menetelmä tarjoaa 11 ihmiseen liittyvää notaatioita.

Taulukko 6.3: Iconclass-luokittelujärjestelmän luokat, joihin aakkosellisessa indeksissä liittyy asiasana ihminen. Kursivoidut määritteet ovat vapaita suomennoksia

11R	ihmiselämä
25FF78	<i>satueläimet ~ alemmat eläinlajit</i>
3	ihminen
34(+98)	ihminen ja eläin
46B2	kaupallisella alalla toimivat henkilöt
46C(+78)	<i>eläinten kuljettaminen ~ ihminen ja eläin</i>
47(+98)	<i>karjaeläimet ~ ihminen ja eläin</i>
58B21	<i>ihmiselämä; 'Vita humana', 'Vita inquieta', 'Vita e l'animo' (Ripa)</i>
59A	<i>ihminen yhteisö</i>
71A31	Jumala muotoilee ihmisen savesta, Aatamin luominen
91E12	<i>Cura muotoilee ihmisen savesta</i>

Menetelmään liittyy kuitenkin ongelmia. Indeksinkään avulla ei löydetä viittauksia kaikkiin ihmistä koskeviin notaatioihin, koska asiasanan ”ihminen” sijaan notaatioon on saatettu valita jokin muu ihmistä kuvaava asiasana kuten ”henkilö”, ”mies” tai ”nainen”. Asiasanojen välillä on ristiviitteitä sanan synonyymeihin ja meronyymeihin, mutta nämäkin viitteet ovat vaillinaisia. Esimerkiksi yhteys ihmisen ja henkilön välillä jää puuttumaan. Asiasanasta ”ihminen” on viittaus vain sanoihin ”rotu”, ”ihmisfiguuri” ja ”ihmiskunta”. Näin ihmistä historiallisena henkilönä merkitsevä notaatio 61B2 jää kokonaan löytymättä, koska tämän notaation yhteydessä ihmiseen viitataan asiasanoilla ”historiallinen henkilö”, ”mies”, ”nainen” ja ”lapsi”.

Teosten kuvailua voidaan pyrkiä rikastamaan etsimällä aiemmin käsitellystä aineistosta teoksia, joissa on käytetty samankaltaista luokittelua ja suosittelemalla tämän perusteella notaatioita teokselle. Menetelmä on vastaava kuin verkkokirjakaupoissa käytetty käyttäjän aiempiin hankintoihin perustuva uusien teosten suosittelujärjestelmä (Fensel, 2004, s. 104). Sen avulla voidaan pyrkiä tunnistamaan teostyyppejä, joissa usein esiintyy samankaltaisia elementtejä. Esimerkiksi asetelmaa kuvaavassa maalauksessa esiintyy usein hedelmiä ja astioita pöydällä, joten mikäli teokselle on

valittu pöytää ja omenaa merkitsevät notaatiot, saattaa suosittelu ehdottaa teokselle myös maljakkoa merkitsevää notaatiota.

6.2.2 Olennaisen sisällön esille nostaminen

Luokittelun tavoitteena on nostaa esille teosten olennainen sisältö. Loppukäyttäjän mielenkiinnonkohteiden selvittämiseksi olisikin tutkittava, millaisia hakuja he tekevät. Hakuja tallentamalla ja analysoimalla voitaisiin selvittää, millä hakusanoilla tulee eniten osumia teokselle valittuihin luokkiin. Näin luokittelijalle voidaan suositella sellaisia notaatioita, jotka esiintyvät usein tiettyyn aihepiiriin kohdistuvissa hauissa.

Sisällön löytymistä voitaisiin myös parantaa kuvaamalla Iconclass-luokittelujärjestelmän notaatiot luvussa 4.3.2 esitettyä tapahtumakehystä käyttäen. Tapahtumakehysten käyttö tukisi ICONCLASS-ontologiaan perustuvaa semanttista päättelyä ja mahdollistaisi kehittyneempien suosittelusääntöjen toteuttamisen.

6.2.3 Luokittelun tehokkuus

Taidemuseoiden kokoelmat ovat laajoja ja siten myös luokittelijalla yhtä teosta kohden käytettävissä oleva aika on rajallinen. Siksi on tärkeää, että suosittelu pystyy tarjoamaan luokittelijalle mahdollisimman hyviä suosituksia, jotta luokittelijan aika ei kulu kaikkien mahdollisten vaihtoehtojen läpikäyntiin.

Suosituksen laatua voidaan arvioida laskemalla suosituksille relevanssipisteytys ja esittämällä luokittelijalle korkeimmat pisteet saaneet suositukset. Relevanssipisteytyksen perusteena voidaan käyttää suositeltavien käsitteiden ontologisia suhteita teoksen luokitteluun jo valittuihin käsitteisiin. Jos käsitteet esimerkiksi viittaavat useisiin samoihin YSO:n käsitteisiin, ne todennäköisesti liittyvät läheisesti toisiinsa.

6.2.4 Hierarkkisen rakenteen hankaluus

Luokittelujärjestelmän hierarkkisen rakenteen muodostamisessa tehdyistä valinnoista johtuen toisiinsa läheisestikin liittyviä käsitteitä kuvaavat notaatiot saattavat jakaantua useisiin eri luokittelujärjestelmän haaroihin, kuten taulukossa 6.3 esitetyt

ihmistä kuvaavat notaatiot. Tästä johtuen sopivien notaatioiden hakeminen luokittelujärjestelmän hierarkkista rakennetta käyttäen voi olla työlästä.

ICONCLASS-ontologian yhdistäminen toisiin ontologioihin, kuten YSO:aan tarjoaa mahdollisuuden hyödyntää myös näiden ontologioiden rakennetta sopivien käsitteiden hakemiseen. Yhteyksiä toisiin ontologioihin voidaan hyödyntää myös lähi-käsitteiden tarjoamiseen etsittäessä sopivaa notaatiota hakusanalla.

6.2.5 Toteutusvaiheet

Luokittelijan apuvälineen prototyypin toteutus on jaettu seuraaviin kolmeen osavaiheeseen, joista ensimmäinen toteutettiin tämän työn yhteydessä. Toinen vaihe on tarkoitus toteuttaa jatkossa FinnONTO-projektin puitteissa. Kolmas vaihe kuvaa edistyneempiä ominaisuuksia, joilla työkalun toimintaa voitaisiin edelleen kehittää.

1. ICONCLASS-ontologian kehittäminen ja yhdistäminen YSO:aan. ICONCLASS-ontologian, asiasanaston ja YSO:n yhdistelmän selaaminen ja haku ONKI-se-laimen avulla.
2. Käyttöliittymän mukauttaminen luokittelutyön tarpeisiin. ICONCLASS-onto-logiaan ja YSO:aan perustuvien suosittelusääntöjen toteuttaminen.
3. Hakujen ja aineiston analysointiin perustuvien kehittyneempien suosittelume-netelmien toteuttaminen. ICONCLASS-ontologian kuvailu tapahtumakehys-ten avulla ja semanttisen päättelyn parantaminen.

7 ARVIOINTI

Tässä luvussa arvioidaan SKOS Core -mallin soveltuvuutta luokittelujärjestelmien ontologisointiin sekä työssä esitetyn ontologiamallin soveltumista Iconclass-järjestelmän kuvaamiseen. Lisäksi arvioidaan ontologian käyttökelpoisuutta sovelluskohteista saatujen kokemusten pohjalta sekä Iconclass-järjestelmän rakennetta kuvaavassa lähdeaineistossa havaittujen puutteiden ja virheiden vaikutusta ontologian muodostamiseen.

7.1 SKOS Core -malli

SKOS Core -malli pyrkii tarjoamaan rakenteen tesaurusten ja luokittelujärjestelmien rakenteen kuvaamiseen. Mallin kehitys on kuitenkin lähtenyt liikkeelle tesaurusten rakenteen kuvaamisesta. Tämä näkyy muun muassa mallin käsitteille määrittämistä ominaisuuksista. Käsitteeseen voidaan liittää sanallisia nimikkeitä, mutta luokittelujärjestelmissä tyypillisten luokittelukoodien esittämiseen ei ole vakiintunutta käytäntöä.

SKOS Core -malli soveltuu rakenteeltaan selkeästi hierarkkisten käsitteistöjen esittämiseen, mutta on riittämätön moniulotteisempien mallien semanttisen rakenteen kuvaamiseen. Iconclass-järjestelmän tapauksessa SKOS Core -mallilla voidaan kuvata luokittelujärjestelmän runko, mutta ei järjestelmän laajennusten rakennetta.

Mallin laajennettavuus antaa mahdollisuuden kehittää SKOS Core -mallin kanssa yhteentoimivia malleja semantiikaltaan monimutkaisemmille ontologioille. Yhteentoimivuuden kannalta haasteen kuitenkin muodostavat useista erillistä osista koostuvat mallit, kuten Iconclass-järjestelmän erilliset avain-hierarkiat, koska tällöin on myös kuvattava mallin osien väliset yhteydet mallin osien sisäisten suhteiden merkitysten täsmentämisen lisäksi.

7.2 ICONCLASS-ontologian malli

Työssä esitetty ICONCLASS-ontologian malli pyrkii kuvaamaan Iconclass-järjestelmää erityisesti loppukäyttäjän ja luokittelijan tarpeista lähtien. Tästä syystä mal-

li toteutettiin SKOS Core -mallia laajentamalla, koska tavoitteena oli Iconclass-järjestelmän rakenteen mahdollisimman yksinkertainen esittäminen köyhdyttämättä kuitenkaan sen rakenteeseen sisältyvää semantiikkaa.

Edellä esitetyistä lähtökohdista johtuen malli ei esitä täydellisesti kaikkia sellaisia rakenteita, joilla on merkitystä Iconclass-järjestelmän ylläpidon kannalta, mutta jotka eivät vaikuta notaation merkityksen tulkintaan. Tähän joukkoon kuuluvat rakenteellinen numerointi -laajennus ja rekursiiviset avaimet, jotka on ontologiassa laajennettu tavallisiksi notaatioiksi ja avaimiksi. Näin ollen esitettyä mallia voidaan käyttää Iconclass-järjestelmää hyödyntävissä haku- ja luokittelusovelluksissa, mutta se ei sovellu järjestelmän ylläpitämiseen ontologiana.

Nimi sulkeissa -laajennuksessa käytössä oleva mahdollisuus vapaaseen asiasanoitukseen tulisi saada mukaan myös ontologisen päättelyn piiriin. Tämän laajennuksen käyttö korostuu erityisesti suomalaisessa aineistossa, jossa järjestelmän valmiista sanastosta ei välttämättä löydy sopivaa asiasanaa. Tämä näkyi myös KulttuuriSampoa varten käsitellyssä aineistossa. Suuri osa tapauksista, joissa suoraa käsitteistä ei löytynyt, liittyi nimi sulkeissa -laajennuksen käyttöön. Ongelmana laajennukseen liittyvän päättelyn parantamisessa ovat kuitenkin kirjavat käytännöt laajennuksen soveltamisessa.

7.3 Ontologian sovellukset

ICONCLASS-ontologia ja sen yhdistäminen YSO:n kanssa mahdollistivat taideteosten kuvailun muuntamisen KulttuuriSammon annotaatioskeemaa noudattavaan yhdenmukaiseen muotoon portaalin muiden aineistotyyppien kanssa. Ontologian avulla teoksiin voitiin myös liittää viitteitä semanttisiin käsitteisiin, joiden välisiä yhteyksiä voidaan hyödyntää järjestelmän suosittelussa ja haussa.

Ontologia avaa uusia mahdollisuuksia myös luokittelun apuvälineen kehittämiseen, kun luokkahierarkian rinnalla ja koneellisen päättelyn apuna voidaan käyttää Iconclass-järjestelmän rakenteen lisäksi myös siihen yhdistettyä YSO:aa. Näin voidaan helpottaa esimerkiksi järjestelmässä hajallaan olevien toisiinsa liittyvien notaatioiden löytymistä. Haasteita luokittelutyökalun kehittämislle muodostavat

Iconclass-luokittelun kannalta keskeisten henkilöiden sekä paikannimien käsittely ja tapahtumia kuvaavien notaatioiden muodostaminen.

7.4 Lähdeaineiston puutteet

Ontologisointityön aikana ilmeni, että Iconclass-järjestelmän rakennetta kuvaava XML-tiedosto sisälsi virheitä luokittelujärjestelmän rakenteiden kuvaamisessa. Luokittelujärjestelmän hierarkista puuttui joitakin haaroja ja avain-laajennuksen käyttöä koskevat tiedot olivat puutteellisia. Nämä ongelmat johtuvat todennäköisesti tiedoston tuottaneen muunnosohjelman virheestä. Nämä puutteet heijastuvat myös ICONCLASS-ontologiaan, mutta niillä ei ole vaikutusta ontologian muodostamisessa käytetyn menetelmän toimivuuteen.

Iconclass-järjestelmän ontologisoinnissa käytetty suomennos oli osittainen, mistä johtuen kaikille notaatioiden määritteille ja asiasanoille ei ollut käytettävissä suomenkielisiä vastineita. Näin myös ontologian käsitteiden vastaavuudet YSO:n käsitteiden kanssa kattavat ICONCLASS-ontologian suomennetuilta osin. Koska aakkosellisen indeksin asiasanat on suomennettu erillään luokittelujärjestelmästä, osalla suomennoksista on sivumerkityksiä, jotka eivät liity niihin luokittelujärjestelmän notaatioihin, joiden kuvaamiseen asiasanaa on käytetty.

Jotta ontologia voitaisiin ottaa tuotantokäyttöön, tulisi työssä kuvattu muunnos suorittaa ajantasaisesta ja luokittelujärjestelmän koko rakenteen esittävästä lähdeaineistosta sekä koko järjestelmän kattavasta suomennoksesta.

8 JOHTOPÄÄTÖKSET

8.1 Vastaukset tutkimuskysymyksiin

Tässä luvussa esitetään johtopäätökset, jotka vastaavat luvussa 1.2 esitettyihin tutkimuskysymyksiin.

1. Millainen Iconclass-järjestelmän rakenne on verrattuna tesarusten ja ontologioiden rakenteeseen?

Iconclass-järjestelmä koostuu hierarkkisesta luokittelujärjestelmästä, aakkosellisesta indeksistä ja bibliografiasta. Tässä työssä näistä käsiteltiin kahta ensimmäistä osaa. Tesaursuksista ja niistä muodostetuista ontologioista poiketen luokittelujärjestelmässä käytetään käsitteiden ensisijaisina nimikkeinä kieliriippumattomia aakkosnumeerisia notaatioita luonnollisella kielellä esitettyjen termien sijaan. Luokittelujärjestelmä on perusrakenteeltaan hierarkkinen jakaantuen kymmeneen pääluokkaan ja niitä tarkentaviin alaluokkiin. Tämänkaltainen hierarkkinen rakenne on tyyppillinen myös ontologioille. Luokittelujärjestelmässä on useita semanttiselta rakenteeltaan monimutkaisia laajennuksia, jotka mahdollistavat asioiden ja tapahtumien yksityiskohtaisen kuvaamisen, mutta joiden semantiikan esittäminen ontologiassa on haasteellista.

Aakkosellisen indeksin tehtävänä on helpottaa sopivien notaatioiden etsimistä luokittelujärjestelmästä tarjoamalla notaatioihin viittauksia sisältävän asiasanaston. Aakkosellinen indeksi on rakenteeltaan tesaursuksen kaltainen ja on tärkeä osa yhdistettäessä ICONCLASS-ontologiaa muihin tesarusten ja asiasanastojen pohjalta muodostettuihin ontologioihin.

2. Mitä ongelmia Iconclass-järjestelmässä on konesemantiikan kannalta?

Luokittelujärjestelmän runko soveltuu hyvin konesemantiikan keinoin mallinnettavaksi. Luokittelujärjestelmästä muodostettu ontologia rakentuu vahvasti hierarkkisen rakenteen pohjalle, sillä luokittelujärjestelmä on rakenteeltaan paikoin hyvin syvä ja sisältää lähinnä hierarkkisia suhteita. Assosiatiivisten ristiviitteiden määrä on vähäinen. Luokittelujärjestelmässä luokkien esittämiseen käytettyjen yksikäsitteis-

ten notaatioiden ansiosta URI-tunnisteiden muodostaminen ontologian käsitteille on suoraviivaista, kunhan URI-muodon rajoitteet huomioidaan ja näille määrittelyä tarvittavat muunnosmenetelmät.

Luokittelujärjestelmän laajennuksista rakenteellinen numerointi ja avain-laajennuksen osana oleva avaimen koostaminen jonosta rekursiivisen avaimen avulla ovat tarkoitettu lähinnä luokittelujärjestelmän ylläpitäjän apuvälineeksi. Nämä laajennukset ovat luokittelijan tai loppukäyttäjän kannalta läpinäkyvättömiä, sillä niiden tulkintaa ei voi tehdä yksin notaatiota lukemalla. Konesemanttisessa käsittelyssä näitä laajennuksia voidaan kuitenkin hyödyntää assosiatiivisten suhteiden etsimisessä luokittelujärjestelmästä.

Luokittelujärjestelmän rungon ja laajennusten välisten suhteiden ilmaiseminen on vaikeaa, sillä laajennuksia voidaan soveltaa vain rajatuissa osissa luokittelujärjestelmää ja laajennuksen merkitys riippuu siitä, missä luokittelujärjestelmän osassa sitä sovelletaan. Tärkein näistä on avain-laajennus, joka määrittelee joukon koodeja, joilla notaatioiden merkitystä voidaan tarkentaa. Tarkennuksen merkitystä ei voida kuitenkaan tulkita yksin laajennuksessa määritellyn koodin perusteella, sillä sen merkitys vaihtelee riippuen siitä, mihin notaatioon se on liitetty.

Luokittelujärjestelmä tarjoaa monessa kohdin mahdollisuuden lisätä luokittelukoodiin sanallisen tarkenteen. Osa näistä tarkenteista on määritelty luokittelujärjestelmässä, mutta monin paikoin on mahdollista käyttää myös vapaamuotoista tarkennetta, joka jää hallitun ontologian ulkopuolelle. Vapaamuotoinen tarkenne mahdollistaa yksityiskohtaisemman lisätiedon tarjoamisen ihmiskäyttäjälle, mutta samalla menetetään luokittelujärjestelmän kieliriippumattomuuden tarjoamat edut.

3. Miten Iconclass-järjestelmän ontologisointi toteutetaan?

Iconclass-luokittelujärjestelmän ontologisointi toteutettiin laajentamalla SKOS Core-mallia sen laajennusohjeiden mukaisesti kahdella uudella luokalla ja yhdellä uudella ominaisuudella siten, että malli soveltuu luokittelujärjestelmän semanttisen rakenteen kuvaamiseen. Haku- ja luokittelukäytössä käyttäjälle läpinäkyvät laajennukset (rakenteellinen numerointi ja rekursiivisen avaimen koostaminen jonosta)

palautettiin ontologiassa perusmuotoihinsa notaatioiksi ja avaimiksi. Aakkosellisen indeksin ontologisointi toteutettiin SKOS Core -mallia käyttäen.

Ontologia muodostettiin XSLT-muunnosten avulla. Tämä menettely soveltui tarkoitukseen hyvin, sillä käytössä oli luokittelujärjestelmän rakenteen ja aakkosellisen indeksin asiasanoituksen kuvaava XML-tiedosto.

4. Miten ICONCLASS-ontologia liitetään yleiseen suomalaiseen ontologiaan (YSO)?

ICONCLASS-ontologia liitettiin YSO:aan hakemalla käsitevastineita luokittelujärjestelmän notaatioihin viittaavien aakkosellisen indeksin avainsanojen ja YSO:n käsitteiden välillä. Menetelmän toimintaa arvioitiin tarkastelemalla käsitevastaavuuksien löytymistä taideteosten annotaatioissa käytetyille notaatioille. Menetelmä toimii hyvin sellaisille notaatioille, joihin ei liity vapaan asiasanoituksen käytön mahdollistavaa nimi sulkeissa -laajennusta.

Vaihtoehtoisena menetelmänä kokeiltiin YSO:n käsitteiden hakua NLP-tekniikan avulla luokittelujärjestelmän notaatioiden määriteteksteistä. Menetelmän käytöstä kuitenkin luovuttiin, koska suomennos sisälsi käännökset vain osalle määriteteksteistä.

5. Miten ICONCLASS-ontologia soveltuu koneelliseen päättelyyn?

ICONCLASS-ontologian aiheenmukainen rakenne tarjoaa mahdollisuuden soveltaa teoksen sisällönkuvailua tulkitsevaa koneellista päättelyä aiheeltaan samankaltaisten teosten etsimiseen ja luokittelua tarkentavien notaatioiden suosittelemiseen teosten luokittelijalle. Ontologian assosiatiivisten suhteiden ja asiasanaston tarjoamien viitteiden avulla voidaan yhdistää myös toisiinsa liittyviä aihepiirejä luokittelujärjestelmän eri osista.

Ontologian yhdistäminen muihin ontologioihin kuten YSO:aan tarjoaa ontologialle vaihtoehtoisen rakenteen, jonka avulla voidaan löytää yhteyksiä sellaisten käsitteiden välille, jotka Iconclass-järjestelmän jaottelussa ovat jakaantuneet luokitteluhierarkian eri haaroihin.

Useat Iconclass-järjestelmän notaatiot kuvaavat erilaisia tapahtumia. Tällaisia ovat

erityisesti avain-laajennusta käyttävät notaatiot. ICONCLASS-ontologian käsitteiden merkityksen kuvaamien tapahtumakehyksillä mahdollistaisi näihin notaatioihin kohdistuvan tarkemman koneellisen päättelyn, kuten tapahtuman osallistujien roolien tunnistamisen. Tämän tyyppistä tapahtumakehystä on käytetty esimerkiksi eräissä KulttuuriSammon aineistotyypeissä. Sen soveltaminen Iconclass-järjestelmään edellyttäisi kuitenkin ontologian käsitteiden kuvailua käsityönä.

8.2 Jatkotutkimuskohteita

Tämän työn puitteissa muodostettiin Iconclass-luokittelujärjestelmästä ICONCLASS-ontologia, kokeiltiin sitä semanttisessa hakusovelluksessa osana KulttuuriSampoja sekä selvitettiin mahdollisuuksia ICONCLASS-ontologiaan perustuvan luokittelijan apuvälineen kehittämiseen.

Yksi semanttisen webin keskeisistä tavoitteista on yhteentoimivuuden mahdollistaminen. Koska Iconclass on kansainvälisesti käytössä oleva järjestelmä, olisi tärkeää, että myös viittaukset järjestelmästä rakennettuun ontologiaan voitaisiin tehdä yksikäsitteisesti. Tämä edellyttää yksimielisyyttä ontologian rakenteesta järjestelmän käyttäjien kesken ja järjestelmää hallitsevan RKD:n hyväksyntää.

ICONCLASS-ontologian tarkempi kuvailu tapahtumakehysten avulla vaikuttaa kiinnostavalta mahdollisuudelta rikastaa ontologiaan perustuvaa semanttista päättelyä. Toteutus vaatii kuitenkin manuaalista työtä, ja tähän liittyvä kysymys on tarjoaako tapahtumakuvaus niin merkittävää etua käsitteivastineiden käyttöön verrattuna, että työhön kannattaa ryhtyä.

Luokittelijan apuvälineen toteutukseen liittyvissä kysymyksissä nousee useassa kohdin keskeiseksi kysymykseksi loppukäyttäjän tarpeisiin vastaaminen. Olemassa olevien Iconclass-järjestelmää hyödyntävien hakujärjestelmien käyttöä on kuitenkin tutkittu vähän, mikä hankaloittaa käyttäjien todellisiin tarpeisiin perustuvien ratkaisujen tekemistä.

8.3 Tulevaisuus

ICONCLASS-ontologiaa hyödyntävien sovellusten kehitystä jatketaan FinnONTO-projektin puitteissa. KulttuuriSampo-portaali tullaan julkaisemaan yleiseen käyttöön vuonna 2007. Iconclass-luokittelijan apuvälineestä toteutetaan prototyyppiversio, jolla pilotoidaan luokittelutyön helpottamista semanttisen suosittelun avulla.

LÄHDELUETTELO

Alvestrand, H. (2001), 'Tags for the Identification of Languages', The Internet Society. RFC 3066. [Viitattu: 13.09.2006] Saatavissa: URL: <<http://www.ietf.org/rfc/rfc3066.txt>>

Anon (2003), 'Verkko-YKL'. [Viitattu: 05.04.2006] Saatavissa: URL: <<http://ykl.kirjastot.fi/>>

Anon (2004), 'Iconclass info: What is Iconclass?'. [Viitattu: 04.04.2006] Saatavissa: URL: <<http://www.iconclass.nl/texts/info01.htm>>

Anon (2005a), 'Iconclass Glossary'. [Viitattu: 09.08.2006] Saatavissa: URL: <<http://www.iconclass.nl/texts/glossary01.htm>>

Anon (2005b), 'Iconclass info: Standardization issues'. [Viitattu: 10.04.2006] Saatavissa: URL: <<http://www.iconclass.nl/texts/info02.htm>>

Anon (2006a), 'ISO 639-2/RA Change Notice - Codes for the representation of names of languages', Library of Congress. [Viitattu: 13.09.2006] Saatavissa: URL: <<http://www.loc.gov/standards/iso639-2/codechanges.html>>

Anon (2006b), 'Muusa-palvelu'. [Viitattu: 16.11.2006] Saatavissa: URL: <<http://www.muusa.net/>>

Anttonen, E. (2003), 'ICONCLASS - Aiheenmukaisen luokitusjärjestelmän suomen-nettu sanasto'. [Viitattu: 29.03.2006] Saatavissa: URL: <<http://www.fng.fi/fng/rootnew/fi/kehys/pdf/ic2002.pdf>>

van Assem, M., Malaise, V., Miles, A. & Schreiber, G. (2006), A Method to Convert Thesauri to SKOS, Teoksessa: 'Proceedings of the Third European Semantic Web Conference (ESWC'06)', Lecture Notes in Computer Science. [Viitattu: 29.03.2006] Saatavissa: URL: <<http://www.cs.vu.nl/~mark/papers/Assem06b.pdf>>

van den Berg, J. (1995), Subject Retrieval in Pictorial Information Systems, Teoksessa: 'Electronic Filing, Registration, and Communication of Visual Historical Data. Abstracts for Round Table no 34 of the 18th International Congress of Historical Sciences', Kööpenhamina, Tanska, ss. 21–29. [Viitattu: 10.05.2006] Saatavissa: URL: <<http://www.iconclass.nl/texts/pub07.htm>>

Berners-Lee, T. (2000), 'Semantic Web on XML', XML 2000. [Viitattu: 26.07.2006] Saatavissa: URL: <<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>>

Berners-Lee, T., Bray, T., Connolly, D., Cotton, P., Fielding, R., Jeckle, M., Lilley, C., Mendelsohn, N., Orchard, D., Walsh, N. & Williams, S. (2004), 'Architecture of the World Wide Web, Volume One'. W3C Recommendation. [Viitattu: 15.05.2006] Saatavissa: URL: <<http://www.w3.org/TR/2004/REC-webarch-20041215/>>

Berners-Lee, T., Fielding, R. T. & Masinter, L. (2005), 'Uniform Resource Identifier (URI): Generic Syntax', The Internet Society. RFC 3986. [Viitattu: 03.05.2006] Saatavissa: URL: <<http://www.gbiv.com/protocols/uri/rfc/rfc3986.html>>

Berners-Lee, T., Hendler, J. & Lassila, O. (2001), 'The Semantic Web', *Scientific American* 284(5), 34–43. [Viitattu: 29.03.2006] Saatavissa: URL: <<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&ref=sciam>>

- Bray, T., Hollander, D. & Layman, A. (toim.) (1999), 'Namespaces in XML'. W3C Recommendation. [Viitattu: 08.05.2006] Saatavissa: URL: <<http://www.w3.org/TR/1999/REC-xml-names-19990114/>>
- Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E. & Yergeau, F. (toim.) (2006), 'Extensible Markup Language (XML) 1.0 (Fourth Edition)'. W3C Recommendation. [Viitattu: 13.09.2006] Saatavissa: URL: <<http://www.w3.org/TR/2006/REC-xml-20060816/>>
- Brickley, D. & Guha, R. V. (toim.) (2004), 'RDF Vocabulary Description Language 1.0: RDF Schema'. W3C Recommendation. [Viitattu: 22.08.2006] Saatavissa: URL: <<http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>>
- Doerr, M. (2003), 'The CIDOC CRM - An Ontological Approach to Semantic Interoperability of Metadata', *AI Magazine* 24(3). [Viitattu: 16.11.2006] Saatavissa: URL: <http://cidoc.ics.forth.gr/docs/ontological_approach.pdf>
- Fensel, D. (2004), *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, 2. painos, Springer-Verlag, Berliini, Saksa. ISBN 3-540-00302-9.
- Fensel, D., Hendler, J., Lieberman, H. & Wahlster, W. (toim.) (2003), *Spinning the Semantic Web. Bringing the World Wide Web to its full potential*, MIT Press, Cambridge, MA, USA. ISBN 0-262-06232-1.
- van Gendt, M., Isaac, A., van der Meij, L. & Schlobach, S. (2006), Semantic Web Techniques for Multiple Views on Heterogeneous Collections: a Case Study, Teoksessa: J. Gonzalo, C. Thanos, M. F. Verdejo & R. C. Carrasco (toim.), 'Proceedings of the 10th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2006)', Vol. 4172 of *Lecture Notes in Computer Science*, Springer Verlag, Alicante, Espanja, ss. 426–437. [Viitattu: 18.10.2006] Saatavissa: URL: <<http://stitch.cs.vu.nl/papers/STITCH-ECDL06.pdf>>
- Gruber, T. R. (1993), 'A Translation Approach to Portable Ontology Specifications', *Knowledge Acquisition* 5(2), 199–220. [Viitattu: 23.05.2006] Saatavissa: URL: <<http://tomgruber.org/writing/ontolingua-kaj-1993.pdf>>
- Heikka, E. (2006), 'Art in Contexts. National Digitization and Content Production Project Town and Again - Images of Urban Finland', Symposium on Digital Semantic Content across Cultures. Louvre, Pariisi, Ranska. [Viitattu: 16.05.2006] Saatavissa: URL: <<http://www.seco.tkk.fi/events/2006/2006-05-04-websemantique/presentations/thursday-1100-town-and-again-paris.pdf>>
- Hendler, J. (2001), 'Agents and the Semantic Web', *IEEE Intelligent Systems* 16(2), 30–37. [Viitattu: 26.07.2006] Saatavissa: URL: <<http://www.etsi2.ugr.es/depar/ccia/mabd/material/adicional/semantic/semantic3.pdf>>
- Hyvönen, E. (2005a), 'Kohti suomalaista semanttista webiä – Suomalaisen semanttisen webin ontologiat (FinnONTO)-hankkeen esittely', FinnONTO-symposium. [Viitattu: 29.03.2006] Saatavissa: URL: <<http://www.seco.tkk.fi/publications/2005/hyvonen-suomalaisen-semanttisen-webin-ontologiat-2005.pdf>>
- Hyvönen, E. (2005b), 'Miksi asiasanastot eivät riitä vaan tarvitaan ontologioita?'. [Viitattu: 29.03.2006] Saatavissa: URL: <<http://www.seco.tkk.fi/publications/2005/hyvonen-miksi-asiasanastot-eivat-riita-2005.pdf>>

Hyvönen, E., Ruotsalo, T., Häggström, T., Salminen, M., Junnila, M., Virkkilä, M., Haaramo, M., Kauppinen, T., Mäkelä, E. & Viljanen, K. (2006), *CultureSampo - Finnish Culture on the Semantic Web: The Vision and First Results*, Teoksessa: 'Semantic Web at Work - Proceedings of the 12th Finnish Artificial Intelligence Conference STeP 2006', Vol. 1, ss. 25–36. [Viitattu: 11.10.2006] Saatavissa: URL: <<http://www.stes.fi/scai2006/proceedings/step2006-25-hyvonen-culturesampo-finnish-culture-on-the-semantic-web.pdf>>

Hyvönen, E., Valo, A., Komulainen, V., Seppälä, K., Kauppinen, T., Ruotsalo, T., Salminen, M. & Ylisalmi, A. (2005), *Finnish National Ontologies for the Semantic Web - Towards a Content and Service Infrastructure*, Teoksessa: 'Proceedings of International Conference on Dublin Core and Metadata Applications (DC 2005)'. [Viitattu: 29.08.2006] Saatavissa: URL: <<http://www.seco.tkk.fi/publications/2005/hyvonen-valo-et-al-finnish-national-2005.pdf>>

Klyne, G. & Carrol, J. J. (toim.) (2004), 'Resource Description Framework (RDF): Concepts and Abstract Syntax'. W3C Recommendation. [Viitattu: 09.05.2006] Saatavissa: URL: <<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>>

Komulainen, V., Valo, A. & Hyvönen, E. (2005), *A Collaborative Ontology Development and Service Framework ONKI*, Teoksessa: 'Proceeding of ESWC 2005, poster papers'. [Viitattu: 10.08.2006] Saatavissa: URL: <<http://www.seco.tkk.fi/publications/2005/komulainen-valo-et-al-a-collaborative-ontology-development-2005.pdf>>

Lowe, E. J. (2005), *Ontology*, Teoksessa: T. Honderich (toim.), 'The Oxford Companion to Philosophy', 2. painos, Oxford University Press. ISBN 0-19-926479-1.

Lüscher, H. (2003), *ICONCLASS - Ensiaskeleet, Raportti, Valtion taidemuseo*. [Viitattu: 12.04.2006] Saatavissa: URL: <<http://www.fng.fi/fng/rootnew/fi/kehys/pdf/icraportti.pdf>>

Lüscher, H. (2004), *ICONCLASS - Ikonografista luokittelua, Pro gradu -työ, Helsingin yliopisto*. [Viitattu: 21.12.2005] Saatavissa: URL: <<http://ethesis.helsinki.fi/julkaisut/hum/taite/pg/luescher/>>

McGuinness, D. L. & van Harmelen, F. (toim.) (2004), 'OWL Web Ontology Language Overview'. W3C Recommendation. [Viitattu: 25.09.2006] Saatavissa: URL: <<http://www.w3.org/TR/2004/REC-owl-features-20040210/>>

Miles, A. (2005a), 'SKOS Core Integrity Testing and Quality Assurance for Instance Data'. Draft. [Viitattu: 23.08.2006] Saatavissa: URL: <<http://isegserv.itd.rl.ac.uk/cvs-public/~checkout~/skos/drafts/integrity.html?rev=1.7>>

Miles, A. (2005b), 'SKOS Extensions'. [Viitattu: 25.09.2006] Saatavissa: URL: <<http://www.w3.org/2004/02/skos/extensions/>>

Miles, A. (2005c), 'SKOS Mapping'. [Viitattu: 25.09.2006] Saatavissa: URL: <<http://www.w3.org/2004/02/skos/mapping/>>

Miles, A. (2006), *SKOS: Requirements for Standardization*, Teoksessa: 'Proceedings of International Conference on Dublin Core and Metadata Applications'. [Viitattu: 25.09.2006] Saatavissa: URL: <<http://isegserv.itd.rl.ac.uk/public/skos/press/dc2006/camera-ready-paper.pdf>>

Miles, A., Baker, T. & Swick, R. (toim.) (2006), 'Best Practice Recipes for Publishing RDF Vocabularies'. W3C Working Draft. [Viitattu: 30.05.2006] Saatavissa: URL: <<http://www.w3.org/TR/2006/WD-swbp-vocab-pub-20060314/>>

Miles, A. & Brickley, D. (toim.) (2005), 'SKOS Core Vocabulary Specification'. W3C Working Draft. [Viitattu: 04.05.2006] Saatavissa: URL: <<http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102/>>

Miles, A. J., Rogers, N. & Beckett, D. (2004), 'Migrating Thesauri to the Semantic Web - Guidelines and case studies for generating RDF encodings of existing thesauri', Deliverable 8.8, SWAD-Europe. [Viitattu: 05.09.2006] Saatavissa: URL: <<http://www.w3.org/2001/sw/Europe/reports/thes/8.8/>>

Miles, A., Matthews, B., Beckett, D., Brickley, D., Wilson, M. & Rogers, N. (2005), SKOS: A language to describe simple knowledge structures for the web, Teoksessa: 'XTech 2005 Conference Proceedings'. [Viitattu: 03.04.2006] Saatavissa: URL: <<http://www.idealliance.org/proceedings/xtech05/papers/03-04-01/>>

Nykänen, O. & Kalliokuusi, V. (1999), Sanastotyön sanastoa, Teoksessa: 'Toimikunnista termitalkoisiin - 25 vuotta sanastotyön asiantuntemusta', Tekniikan sanastokeskus. [Viitattu: 20.04.2006] Saatavissa: URL: <<http://www.tsk.fi/fi/san/sts.html>>

Poole, D., Mackworth, A. & Goebel, R. (1998), *Computational Intelligence: A Logical Approach*, Oxford University Press, New York, NY, USA. ISBN 0-19-510270-3.

Salminen, M. (2006), Kuvien ja videoiden semanttinen sisällönkuvailu, Pro gradu -työ, Helsingin yliopisto. [Viitattu: 14.06.2006] Saatavissa: URL: <<http://www.seco.tkk.fi/publications/2006/salminen-sisallonkuvailu-2006.pdf>>

Svensson, L. (2006), 'How to use notations from classification schemes in SKOS'. SWAD-Europe project's work on RDF and Thesaurus systems mailing list public-esw-thes@w3.org. [Viitattu: 22.08.2006] Saatavissa: URL: <<http://lists.w3.org/Archives/Public/public-esw-thes/2006Feb/0029.html>>

Vehviläinen, A., Alm, O. & Hyvönen, E. (2006), 'Combining Case-Based Reasoning and Semantic Indexing in a Question-Answer Service'. Poster paper, 1st Asian Semantic Web Conference (ASWC2006). [Viitattu: 26.07.2006] Saatavissa: URL: <<http://www.seco.tkk.fi/publications/2006/vehvilainen-alm-hyvonen-combining-cbr-and-semantic-indexing-in-QA-service.pdf>>

Viljanen, K. (2006), Monilähteinen suosittelu semanttisessa webissä, Pro gradu -työ, Helsingin yliopisto. [Viitattu: 16.11.2006] Saatavissa: URL: <<http://www.seco.tkk.fi/publications/2006/viljanen-monilahteinen-suosittelu-2006.pdf>>

Voss, J. (toim.) (2006), 'Quick Guide to Publishing a Classification Scheme on the Semantic Web'. [Viitattu: 13.09.2006] Saatavissa: URL: <<http://esw.w3.org/topic/SkosDev/ClassificationPubGuide?action=recall&rev=10>>

Wielemaker, J., Schreiber, A. T. & Wielinga, B. J. (2003), Supporting Semantic Image Annotation and Search, Teoksessa: S. Handschuh & S. Staab (toim.), 'Annotation for the Semantic Web', Vol. 96 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, Amsterdam, Alankomaat, ss. 147–155. [Viitattu: 10.05.2006] Saatavissa: URL: <<http://www.cs.vu.nl/~guus/papers/Wielemaker02a.pdf>>

Wielinga, B. J., Schreiber, A. T., Wielemaker, J. & Sandberg, J. A. C. (2001), From thesaurus to ontology, Teoksessa: 'K-CAP '01: Proceedings of the 1st international conference on Knowledge capture', ACM Press, New York, NY, USA, ss. 194–201. [Viitattu: 21.12.2005] Saatavissa: URL: <<http://www.cs.vu.nl/~guus/papers/Wielinga01a.pdf>>

LIITE 1. ICONCLASS-LAAJENNUS SKOS CORE -MALLIIN

Alla on ICONCLASS-ontologian esittämistä varten toteutetun SKOS Core -mallin laajennuksen määrittely esitettynä Turtle RDF -syntaksia käyttäen.

```
#####
# This file declares an Iconclass extension to the SKOS Core Vocabulary.
#
# This file uses Turtle RDF syntax, see:
# http://www.dajobe.org/2004/01/turtle/

# Namespace prefixes
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix vs: <http://www.w3.org/2003/06/sw-vocab-status/ns#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema-datatypes> .
@prefix ice: <http://seco.tkk.fi/onto/iconclass/iconclass-skos-ext#> .

# Classes
ice:NotationConcept a rdfs:Class;
    rdfs:label 'Notation Concept'@en;
    skos:definition 'An Iconclass notation element.'@en;
    rdfs:subClassOf skos:Concept;
    vs:term_status 'unstable';
    dcterms:issued '2006-06-07'^^xsd:date;
    dcterms:modified '2006-09-04'^^xsd:date;
.

ice:KeyConcept a rdfs:Class;
    rdfs:label 'Key Concept'@en;
    skos:definition 'An Iconclass key element defined for a certain
        range of notations.'@en;
    rdfs:subClassOf skos:Concept;
    vs:term_status 'unstable';
    dcterms:issued '2006-01-11'^^xsd:date;
    dcterms:modified '2006-09-04'^^xsd:date;
.

# Properties
ice:hasKeyConcept a rdf:Property;
    rdfs:label 'has a specified key'@en;
    skos:definition 'The concept has a specified Iconclass key element
        describing the concept.'@en;
    rdfs:subPropertyOf skos:semanticRelation;
    rdfs:domain ice:NotationConcept;
    rdfs:range ice:KeyConcept;
    vs:term_status 'unstable';
```

dcterms:issued '2006-01-11'^^xsd:date;
dcterms:modified '2006-09-04'^^xsd:date;

LIITE 2. MAALAUKSEN KULLERVON KIROUS ICONCLASS-ANNOTAATIO

Alla on esimerkki kuvassa 6.1 esitettyyn maalaukseen Kullervon kirous liittyvän Iconclass-annotaatiosta KulttuuriSammossa. Luettavuuden parantamiseksi annotaatio on esitetty Turtle RDF -syntaksia käyttäen. KulttuuriSammossa annotaatiot on tallennettu Jena-tietokantamalliin.

```
@prefix ann:      <http://kulttuurisampo.fi/annotaatio#> .
@prefix ic:       <http://seco.tkk.fi/onto/iconclass/iconclass#> .
@prefix ice:      <http://seco.tkk.fi/onto/iconclass/iconclass-skos-ext#> .
@prefix tilanne:  <http://yso.fi/tilanne#> .
```

```
ann:Instance-maalaus-at-A_II_765-tilanne-1161007572473
```

```
  a tilanne:tilanne ;
```

```
  tilanne:mapIC
```

```
    ann:not_81__KALEVALA-1161007579697 ,
```

```
    ann:not_82A_____1161007581122 ,
```

```
    ann:not_13B42-1161007574125 ,
```

```
    ann:not_56E2-1161007579276 ,
```

```
    ann:not_31D112-1161007578708 ,
```

```
    ann:not_31A231-1161007576799 ,
```

```
    ann:not_31A25113-1161007577365 ,
```

```
    ann:not_34B11-1161007578992 ,
```

```
    ann:not_25H15-1161007576418 ,
```

```
    ann:not_25H1124-1161007576130 ,
```

```
    ann:not_25G3_____1161007575739 ,
```

```
    ann:not_25G1-1161007574839
```

```
ann:not_81__KALEVALA-1161007579697 a ic:not_81__KALEVALA .
```

```
ann:not_82A_____1161007581122 a ic:not_82A_____ .
```

```
ann:not_13B42-1161007574125 a ic:not_13B42 .
```

```
ann:not_56E2-1161007579276 a ic:not_56E2 .
```

```
ann:not_31D112-1161007578708 a ic:not_31D112 .
```

```
ann:not_31A231-1161007576799 a ic:not_31A231 .
```

```
ann:not_31A25113-1161007577365 a ic:not_31A25113 .
```

```
ann:not_34B11-1161007578992 a ic:not_34B11 .
```

```
ann:not_25H15-1161007576418 a ic:not_25H15 .
```

```
ann:not_25H1124-1161007576130 a ic:not_25H1124 .
```

```
ann:not_25G3_____1161007575739 a ic:not_25G3 .
```

```
ann:not_25G1-1161007574839
```

```
  a ic:not_25G1 ;
```

```
  ice:hasKeyConcept ic:key_25G__21
```